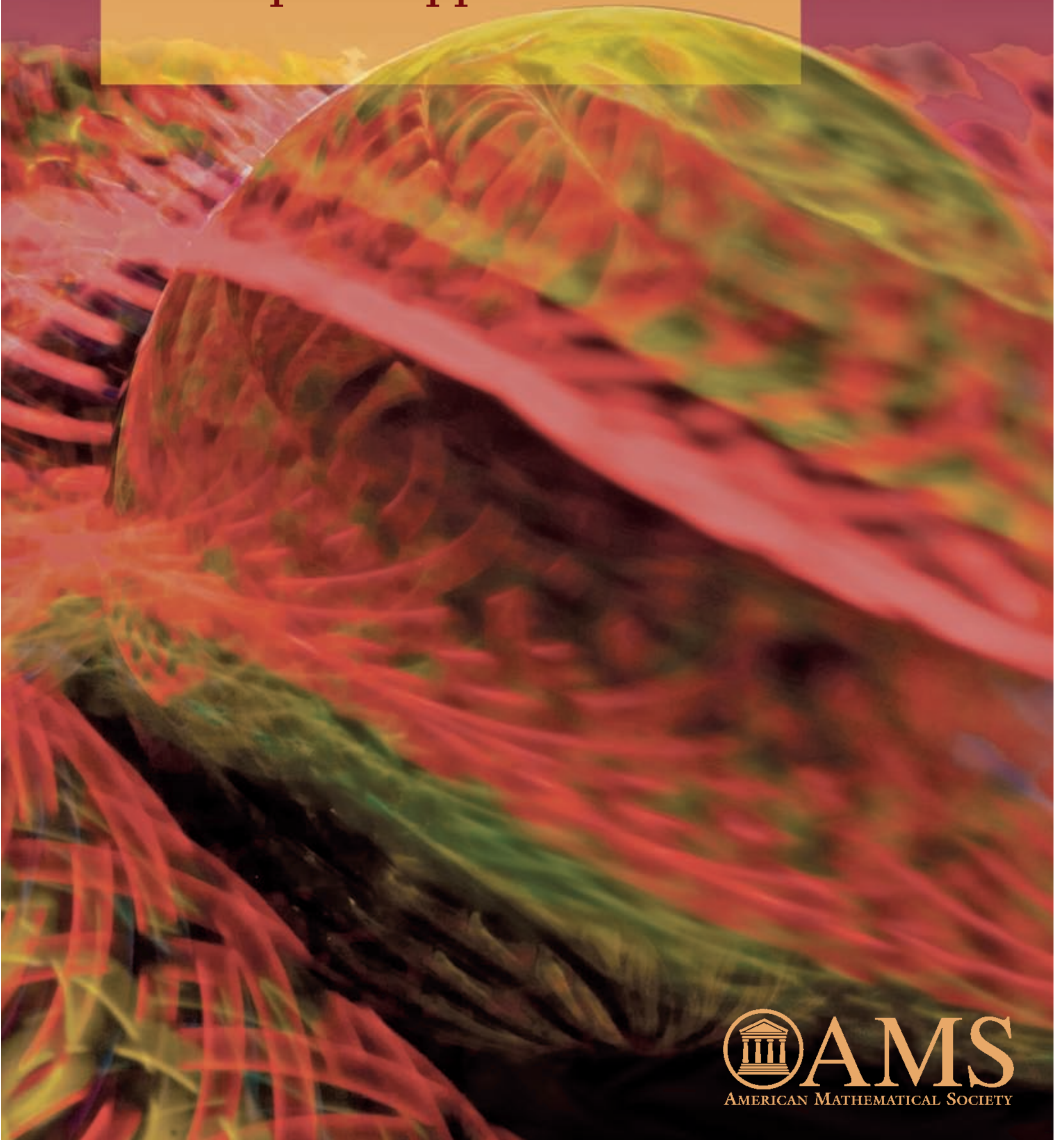
STUDENT MATHEMATICAL LIBRARY
Volume 79

# Matrix Groups for Undergraduates

## Second Edition

Kristopher Tapp



AMS
AMERICAN MATHEMATICAL SOCIETY

# Matrix Groups for Undergraduates

## Second Edition

STUDENT MATHEMATICAL LIBRARY
Volume 79

# Matrix Groups for Undergraduates

## Second Edition

Kristopher Tapp

AMERICAN MATHEMATICAL SOCIETY
Providence, Rhode Island

2010 *Mathematics Subject Classification.* Primary 20-02, 20G20; Secondary 20C05, 22E15.

For additional information and updates on this book, visit
**www.ams.org/bookpages/stml-79**

**Library of Congress Cataloging-in-Publication Data**

Names: Tapp, Kristopher, 1971–

Title: Matrix groups for undergraduates / Kristopher Tapp.

Description: Second edition. — Providence, Rhode Island : American Mathematical Society, [2016] — Series: Student mathematical library ; volume 79 — Includes bibliographical references and index.

Identifiers: LCCN 2015038141 — ISBN 9781470427221 (alk. paper)

Subjects: LCSH: Matrix groups. — Linear algebraic groups. — Compact groups. — Lie groups. — AMS: Group theory and generalizations – Research exposition (monographs, survey articles). msc — Group theory and generalizations – Linear algebraic groups and related topics – Linear algebraic groups over the reals, the complexes, the quaternions. msc — Group theory and generalizations – Representation theory of groups – Group rings of finite groups and their modules. msc — Topological groups, Lie groups – Lie groups – General properties and structure of real Lie groups. msc

Classification: LCC QA184.2 .T37 2016 — DDC 512/.2–dc23 LC record available at http://lccn.loc.gov/2015038141



∞ The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1　　21 20 19 18 17 16

# Contents

# Why study matrix groups?

A **matrix group** means a group of invertible matrices. This definition sounds simple enough and purely algebraic. You know from linear algebra that invertible matrices represent geometric motions (i.e., linear transformations) of vector spaces, so maybe it's not so surprising that matrix groups are useful within geometry. It turns out that matrix groups pop up in virtually any investigation of objects with symmetries, such as molecules in chemistry, particles in physics, and projective spaces in geometry. Here are some examples of how amazingly ubiquitous matrix groups have become in mathematics, physics and other fields:

- **Four-dimensional topology**, **particle physics** and **Yang-Mills connections** are inter-related theories based heavily on matrix groups, particularly on a certain double-cover between two matrix groups (see Section 7 of Chapter 8).
- **Graphics programmers** use matrix groups for rotating and translating three-dimensional objects on a computer screen (see Section 6 of Chapter 3).

- The theory of **differential equations** relies on matrix groups, particularly on matrix exponentiation (see Chapter 6).

- The **shape of the universe** might be a quotient of a certain matrix group, $Sp(1)$, as recently proposed by Jeff Weeks (see Section 6 of Chapter 8). Weeks writes, "Matrix groups model possible shapes for the universe. Conceptually one thinks of the universe as a single multi-connected space, but when cosmologists roll up their sleeves to work on such models they find it far easier to represent them as a simply connected space under the action of a matrix group."

- **Quantum computing** is based on the group of unitary matrices (see Section 2 of Chapter 3). William Wootters writes, "A quantum computation, according to one widely used model, is nothing but a sequence of unitary transformations. One starts with a small repertoire of simple unitary matrices, some $2 \times 2$ and some $4 \times 4$, and combines them to generate, with arbitrarily high precision, an approximation to any desired unitary transformation on a huge vector space."

- In a **linear algebra** course, you may have learned that certain types of matrices can be diagonalized or put into other nice forms. The theory of matrix groups provides a beautifully uniform way of understanding such normal forms (see Chapter 9), which are essential tools in disciplines ranging from topology and geometry to discrete math and statistics.

- **Riemannian geometry** relies heavily on matrix groups, in part because the isometry group of any compact Riemannian manifold is a matrix group. More generally, since the work of Klein, the word "geometry" itself is often understood as the study of invariants of the action of a matrix group on a space.

Matrix groups are used in algebraic geometry, complex analysis, group and ring theory, number theory, quantum physics, Einstein's special relativity, Heisenberg's uncertainty principle, quark theory, Fourier series, combinatorics, and many more areas; see Howe's article [**10**]. Howe writes that matrix groups "touch a tremendous spectrum of mathematical areas...the applications are astonishing in their pervasiveness and sometimes in their unexpectedness."

You will discover that matrix groups are simultaneously algebraic and geometric objects. This text will help you build bridges between your knowledge of algebra and geometry. In fact, the beautiful richness of the subject derives from the interplay between the algebraic and geometric structure of matrix groups. You'll see.

My goal is to develop rigorously and clearly the basic structures of matrix groups. This text is elementary, requires few prerequisites, and provides substantial geometric motivation. Whenever possible, my approach is concrete and driven by examples. Exploring the symmetries of a sphere is a motivating thread woven through the text. You will need only the following prerequisites:

- **Calculus:** topics through multivariable calculus, with a brief introduction to complex numbers including Euler's formula
$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$
- **Linear Algebra:** determinant, trace, eigenvalues, eigenvectors, vector spaces, linear transformations and their relationship to matrices, change of basis via conjugation.
- **Abstract Algebra:** groups, normal subgroups, quotient groups, abelian groups, fields.
- **Analysis (optional):** topology of Euclidean space (open, closed, limit point, compact, connected), sequences and series, continuous and differentiable functions from $\mathbb{R}^m$ to $\mathbb{R}^n$, the inverse function theorem.

The analysis prerequisites are optional. I will develop these analysis topics from scratch for readers seeing this material for the first time, but since this is not an analysis textbook, I will not feel obliged to include complete proofs of analysis theorems.

I believe that matrix groups should become a more common staple of the undergraduate curriculum; my hope is that this text will help allow a movement in that direction.

I am indebted to several authors of previous texts about matrix groups, particularly Curtis [**3**], Howe [**10**], Baker [**1**], Rossmann [**11**] and Hall [**7**]. I wish to thank Charity for support, love and understanding as I wrote this book. Finally, I wish to dedicate this text to Willow Jean Tapp, born March 17, 2004.

Chapter 1

# Matrices

In this chapter, we define quaternionic numbers and discuss basic algebraic properties of matrices, including the correspondence between matrices and linear transformations. We begin with a visual example that motivates the topic of matrix groups.

## 1. Rigid motions of the sphere: a motivating example

The simplest interesting matrix group, called $SO(3)$, can be described in the following (admittedly imprecise) way:

$$SO(3) = \text{all positions of a globe on a fixed stand.}$$

Three elements of $SO(3)$ are pictured in Figure 1. Though the globe always occupies the same place in space, the three elements differ in the directions where various countries face.

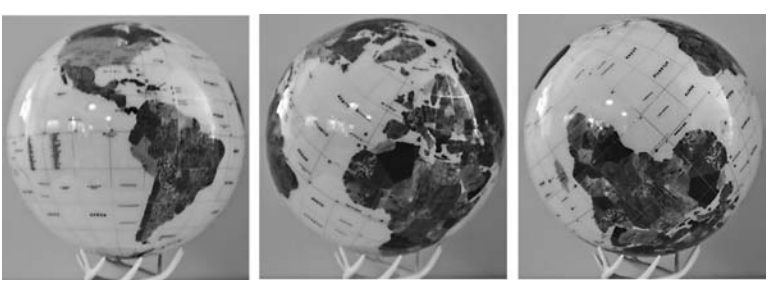


**Figure 1.** Three elements of $SO(3)$.

Let's (arbitrarily) call the first picture "the identity". Every other element of $SO(3)$ is achieved, starting with the identity, by physically moving the globe in some way. $SO(3)$ becomes a group under composition of motions (since different motions might place the globe in the same position, think about why this group operation is well-defined). Several questions come to mind.

**Question 1.1.** *Is $SO(3)$ an abelian group?*

The North Pole of the globe faces up in the identity position. Rotating the globe around the axis through the North Pole provides a "circle's worth" of elements of $SO(3)$ for which the North Pole faces up. Similarly, there is a circle's worth of elements of $SO(3)$ for which the North Pole is located as in picture 2, or at any other point of the globe. Any element of $SO(3)$ is achieved, starting with the identity, by first moving the North Pole to the correct position and then rotating about the axis through its new position. It is therefore natural to ask:

**Question 1.2.** *Is there a natural bijection between $SO(3)$ and the product $S^2 \times S^1 = \{(p, \theta) \mid p \in S^2, \theta \in S^1\}$?*

Here $S^2$ denotes the sphere (the surface of the globe) and $S^1$ denotes the circle, both special cases of the general definition of an **n-dimensional sphere**:

$$S^n = \{(x_1, ..., x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1^2 + \cdots + x_{n+1}^2 = 1\}.$$

Graphics programmers, who model objects moving and spinning in space, need an efficient way to represent the rotation of such objects. A bijection between $SO(3)$ and $S^2 \times S^1$ would help, allowing any rotation to be coded using only three real numbers – two that locate a point of $S^2$ (latitude and longitude) and one angle that locates a point of $S^1$. If no such bijection exists, can we nevertheless understand the shape of $SO(3)$ sufficiently well to somehow parameterize its elements via three real numbers?

One is tempted to refer to elements of $SO(3)$ as "rotations" of the sphere, but perhaps there are motions more complicated than rotations.

**Question 1.3.** *Can every element of $SO(3)$ be achieved, starting with the identity, by rotating through some angle about some single axis?*

If so, then for any element of $SO(3)$, there must be a pair of antipodal points of the globe in their identity position.

You might borrow your roommate's basketball and use visual intuition to guess the correct answers to Questions 1.1, 1.2 and 1.3. But our definition of $SO(3)$ is probably too imprecise to lead to rigorous proofs of your answers. We will return to these questions after developing the algebraic background needed to define $SO(3)$ in a more precise way, as a group of matrices.

## 2. Fields and skew-fields

A matrix is an array of numbers, but what type of numbers? Matrices of real numbers and matrices of complex numbers are familiar. Are there other good choices? We need to add, multiply and invert matrices, so we must choose a number system with a notion of addition, multiplication, and division; in other words, we must choose a field or a skew-field.

**Definition 1.4.** *A* ***skew-field*** *is a set, $\mathbb{K}$, together with operations called addition (denoted "$+$") and multiplication (denoted "$\cdot$") satisfying:*

(1) $a \cdot (b + c) = a \cdot b + a \cdot c$ *and* $(b + c) \cdot a = b \cdot a + c \cdot a$.

(2) $\mathbb{K}$ *is an abelian group under addition, with identity denoted as "0".*

(3) $\mathbb{K} - \{0\}$ *is a group under multiplication, with identity denoted as "1".*

*A skew-field in which multiplication is commutative ($a \cdot b = b \cdot a$ for all $a, b \in \mathbb{K}$) is called a* ***field****.*

The real numbers, $\mathbb{R}$, and the rational numbers, $\mathbb{Q}$, are fields. The plane $\mathbb{R}^2$ is NOT a field under the operations of component-wise addition and multiplication:

$$(a, b) + (c, d) = (a + c, b + d)$$
$$(a, b) \cdot (c, d) = (ac, bd),$$

because, for example, the element $(5, 0)$ does not have a multiplicative inverse; that is, no element times $(5, 0)$ equals $(1, 1)$, which is the only possible multiplicative identity element. A similar argument shows that for $n > 1$, $\mathbb{R}^n$ is not a field under component-wise addition and multiplication.

In order to make $\mathbb{R}^2$ into a field, we use component-wise addition, but we define a more clever multiplication operation:

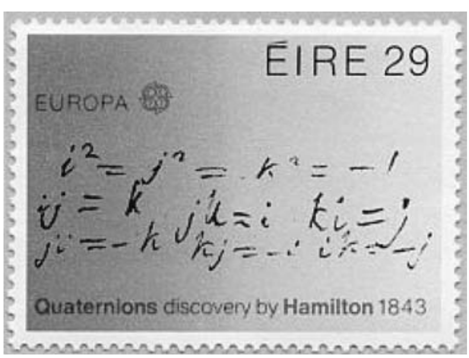
$$(a, b) \cdot (c, d) = (ac - bd, ad + bc).$$

If we denote $(a, b) \in \mathbb{R}^2$ symbolically as $a + b\mathbf{i}$, then this multiplication operation becomes familiar complex multiplication:

$$(a + b\mathbf{i}) \cdot (c + d\mathbf{i}) = (ac - bd) + (ad + bc)\mathbf{i}.$$

It is straightforward to check that $\mathbb{R}^2$ is a field under these operations; it is usually denoted $\mathbb{C}$ and called **the complex numbers**.

## 3. The quaternions




Is it possible to contrive a multiplication operation which, together with component-wise addition, makes $\mathbb{R}^n$ into a skew-field for $n > 2$? This is an important and difficult question. In 1843 Hamilton discovered that the answer is yes for $n = 4$.

To describe this multiplication rule, we will denote an element $(a,b,c,d) \in \mathbb{R}^4$ symbolically as $a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$. We then define a multiplication rule for the symbols $\{1, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$. The symbol "1" acts as expected:

$$\mathbf{i} \cdot 1 = 1 \cdot \mathbf{i} = \mathbf{i}, \qquad \mathbf{j} \cdot 1 = 1 \cdot \mathbf{j} = \mathbf{j} \qquad \mathbf{k} \cdot 1 = 1 \cdot \mathbf{k} = \mathbf{k}.$$

The other three symbols square to $-1$:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1.$$

Finally, the product of two of $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ equals plus or minus the third:

$$\begin{array}{ccc} \mathbf{i} \cdot \mathbf{j} = \mathbf{k}, & \mathbf{j} \cdot \mathbf{k} = \mathbf{i}, & \mathbf{k} \cdot \mathbf{i} = \mathbf{j}, \\ \mathbf{j} \cdot \mathbf{i} = -\mathbf{k}, & \mathbf{k} \cdot \mathbf{j} = -\mathbf{i}, & \mathbf{i} \cdot \mathbf{k} = -\mathbf{j}. \end{array}$$

This sign convention can be remembered using Figure 2.

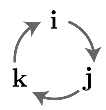



**Figure 2.** The quaternionic multiplication rule.

This multiplication rule for $\{1, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$ extends linearly to a multiplication on all of $\mathbb{R}^4$. For example,

$$\begin{aligned} (2 + 3\mathbf{k}) \cdot (\mathbf{i} + 7\mathbf{j}) &= 2\mathbf{i} + 14\mathbf{j} + 3\mathbf{k}\mathbf{i} + 21\mathbf{k}\mathbf{j} \\ &= 2\mathbf{i} + 14\mathbf{j} + 3\mathbf{j} - 21\mathbf{i} \\ &= -19\mathbf{i} + 17\mathbf{j}. \end{aligned}$$

The product of two arbitrary elements has the following formula:

$$\begin{aligned} (1.1) \quad & (a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) \cdot (x + y\mathbf{i} + z\mathbf{j} + w\mathbf{k}) \\ &= (ax - by - cz - dw) + (ay + bx + cw - dz)\mathbf{i} \\ &\quad + (az + cx + dy - bw)\mathbf{j} + (aw + dx + bz - cy)\mathbf{k}. \end{aligned}$$

The set $\mathbb{R}^4$, together with component-wise addition and the above-described multiplication operation, is denoted as $\mathbb{H}$ and called the **quaternions**. The quaternions have proven to be fundamental in several areas of math and physics. They are almost as important and as natural as the real and complex numbers.

To prove that $\mathbb{H}$ is a skew-field, the only difficult step is verifying that every non-zero element has a multiplicative inverse. For this, it is useful to define the **conjugate** and the **norm** of an arbitrary element $q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \in \mathbb{H}$ as follows:

$$\bar{q} = a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}$$
$$|q| = \sqrt{a^2 + b^2 + c^2 + d^2}.$$

It is straightforward to check that $q \cdot \bar{q} = \bar{q} \cdot q = |q|^2$ and therefore that $\frac{\bar{q}}{|q|^2}$ is a multiplicative inverse of $q$.

The rule for multiplying two quaternions with no $\mathbf{j}$ or $\mathbf{k}$ components agrees with our multiplication rule in $\mathbb{C}$. We therefore have skew-field inclusions:

$$\mathbb{R} \subset \mathbb{C} \subset \mathbb{H}.$$

Any real number commutes with every element of $\mathbb{H}$. In Exercise 1.18, you will show that only real numbers have this property. In particular, every non-real complex number fails to commute with some elements of $\mathbb{H}$.

Any complex number can be expressed as $z = a + b\mathbf{i}$ for some $a, b \in \mathbb{R}$. Similarly, any quaternion can be expressed as $q = z + w\mathbf{j}$ for some $z, w \in \mathbb{C}$, since:

$$a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} = (a + b\mathbf{i}) + (c + d\mathbf{i})\mathbf{j}.$$

This analogy between $\mathbb{R} \subset \mathbb{C}$ and $\mathbb{C} \subset \mathbb{H}$ is often useful.

In this book, the elements of matrices are always either real, complex, or quaternionic numbers. Other fields, like $\mathbb{Q}$ or the finite fields, are used in other branches of mathematics but for our purposes would lead to a theory of matrices with insufficient geometric structure. We want groups of matrices to have algebraic and geometric properties, so we restrict to skew-fields that look like $\mathbb{R}^n$ for some $n$. This way, groups of matrices are subsets of Euclidean spaces and therefore inherit geometric notions like distances and tangent vectors.

But is there a multiplication rule that makes $\mathbb{R}^n$ into a skew-field for values of $n$ other than $1, 2$ and $4$? Do other (substantially different) multiplication rules for $\mathbb{R}^1, \mathbb{R}^2$ and $\mathbb{R}^4$ exist? Can $\mathbb{R}^4$ be made into a field rather than just a skew-field? The answer to all of these questions is NO. More precisely, Frobenius proved in 1877 that

$\mathbb{R}, \mathbb{C}$ and $\mathbb{H}$ are the only associative real division algebras, up to the natural notion of equivalence [**4**].

**Definition 1.5.** *An associative* ***real division algebra*** *is a real vector space, $\mathbb{K}$, with a multiplication rule, that is a skew-field under vector-addition and multiplication, such that for all $a \in \mathbb{R}$ and all $q_1, q_2 \in \mathbb{K}$:*

$$a(q_1 \cdot q_2) = (aq_1) \cdot q_2 = q_1 \cdot (aq_2).$$

The final hypothesis relates multiplication and scalar multiplication. It insures that $\mathbb{K}$ has a subfield isomorphic to $\mathbb{R}$, namely, all scalar multiples of the multiplicative identity 1.

We will not prove Frobenius' theorem; we require it only for reassurance that we are not omitting any important number systems from our discussion. There is an important multiplication rule for $\mathbb{R}^8$, called **octonian** multiplication, but it is not associative, so it makes $\mathbb{R}^8$ into something weaker than a skew-field. We will not consider the octonians.

*In this book, $\mathbb{K}$ always denotes one of $\{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$, except where stated otherwise.*

## 4. Matrix operations

In this section, we briefly review basic notation and properties of matrices. Let $M_{m,n}(\mathbb{K})$ denote the set of all $m$ by $n$ matrices with entries in $\mathbb{K}$. For example,

$$M_{2,3}(\mathbb{C}) = \left\{ \begin{pmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \end{pmatrix} \middle| z_{ij} \in \mathbb{C} \right\}.$$

Denote the space $M_{n,n}(\mathbb{K})$ of **square matrices** as simply $M_n(\mathbb{K})$. If $A \in M_{m,n}(\mathbb{K})$, then $A_{ij}$ denotes the element in row $i$ and column $j$ of $A$.

Addition of same-dimension matrices is defined component-wise, so that

$$(A + B)_{ij} = A_{ij} + B_{ij}.$$

The product of $A \in M_{m,n}(\mathbb{K})$ and $B \in M_{n,l}(\mathbb{K})$ is the element $AB \in M_{m,l}(\mathbb{K})$ defined by the familiar formula:

$$(AB)_{ij} = (\text{row } i \text{ of } A) \cdot (\text{column } j \text{ of } B) = \sum_{s=1}^{n} A_{is}B_{sj}. \tag{1.2}$$

Matrix multiplication is not generally commutative.

Denote a **diagonal matrix** as in this example:

$$\text{diag}(1,2,3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

The **identity matrix** is:

$$I = \text{diag}(1, ..., 1).$$

The **transpose** of $A \in M_{m,n}(\mathbb{K})$ is the matrix $A^T \in M_{n,m}$ obtained by interchanging the rows and columns of $A$, so that:

$$(A^T)_{ij} = A_{ji}.$$

For example,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

It is straightforward to check that if $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ then

$$(A \cdot B)^T = B^T \cdot A^T \tag{1.3}$$

for any matrices $A$ and $B$ of compatible dimensions to be multiplied.

Matrix multiplication and addition interact as follows:

**Proposition 1.6.** *For all $A, B, C \in M_n(\mathbb{K})$,*

(1) $A \cdot (B \cdot C) = (A \cdot B) \cdot C$.

(2) $(A + B) \cdot C = A \cdot C + B \cdot C$ *and* $C \cdot (A + B) = C \cdot A + C \cdot B$.

(3) $A \cdot I = I \cdot A = A$.

The **trace** of a square matrix $A \in M_n(\mathbb{K})$ is defined as the sum of its diagonal entries:

$$\text{trace}(A) = A_{11} + \cdots + A_{nn}.$$

When $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, we have the familiar property for $A, B \in M_n(\mathbb{K})$:

$$\operatorname{trace}(AB) = \operatorname{trace}(BA). \tag{1.4}$$

Since multiplication in $\mathbb{H}$ is not commutative, this property is false even in $M_1(\mathbb{H})$.

When $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, the **determinant** function,

$$\det : M_n(\mathbb{K}) \to \mathbb{K},$$

is familiar. It can be defined recursively by declaring that the determinant of $A \in M_1(\mathbb{K})$ equals its single element, and the determinant of $A \in M_{n+1}(\mathbb{K})$ is defined in terms of determinants of elements of $M_n(\mathbb{K})$ by the expansion of minors formula:

$$\det(A) = \sum_{j=1}^{n+1} (-1)^{j+1} \cdot A_{1j} \cdot \det(A[1,j]), \tag{1.5}$$

where $A[i,j] \in M_n(\mathbb{K})$ is the matrix obtained by crossing out row $i$ and column $j$ from $A$. For example,

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} [2,1] = \begin{pmatrix} b & c \\ h & i \end{pmatrix}.$$

Thus, the determinant of a $3 \times 3$ matrix is:

$$\begin{aligned} \det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} &= a \cdot \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \cdot \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} \\ &\quad + c \cdot \det \begin{pmatrix} d & e \\ g & h \end{pmatrix} \\ &= a(ei - fh) - b(di - fg) + c(dh - eg) \\ &= aei + bfg + cdh - (afh + bdi + ceg). \end{aligned}$$

It is clear that $\det(I) = 1$. In a linear algebra course, one proves that for all $A, B \in M_n(\mathbb{K})$,

$$\det(A \cdot B) = \det(A) \cdot \det(B). \tag{1.6}$$

We postpone defining the determinant of a quaternionic matrix until the next chapter. Exercise 1.5 at the end of this chapter demonstrates why Equation 1.5 is insufficient when $\mathbb{K} = \mathbb{H}$.

Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$. When $a \in \mathbb{K}$ and $A \in M_{n,m}(\mathbb{K})$, we define $a \cdot A \in M_{n,m}(\mathbb{K})$ to be the result of left-multiplying the elements of $A$ by $a$:

$$(a \cdot A)_{ij} = a \cdot A_{ij}.$$

This operation is called **left scalar multiplication**. The operations of matrix addition and left scalar multiplication make $M_{n,m}(\mathbb{K})$ into a *left vector space* over $\mathbb{K}$.

**Definition 1.7.** *A* ***left vector space*** *over a skew-field $\mathbb{K}$ is a set $M$ with an addition operation from $M \times M$ to $M$ (denoted $A, B \mapsto A+B$) and scalar multiplication operation from $\mathbb{K} \times M$ to $M$ (denoted $a, A \mapsto a \cdot A$) such that $M$ is an abelian group under addition, and for all $a, b \in \mathbb{K}$ and all $A, B \in M$,*

(1) $a \cdot (b \cdot A) = (a \cdot b) \cdot A.$

(2) $1 \cdot A = A.$

(3) $(a + b) \cdot A = a \cdot A + b \cdot A.$

(4) $a \cdot (A + B) = a \cdot A + a \cdot B.$

This exactly matches the familiar definition of a vector space. Familiar terminology for vector spaces over fields, like **subspaces**, **bases**, **linear independence**, and **dimension**, make sense for left vector spaces over skew-fields. For example:

**Definition 1.8.** *A subset $W$ of a left vector space $V$ over a skew-field $\mathbb{K}$ is called a $\mathbb{K}$-**subspace** (or just a **subspace**) if for all $a, b \in \mathbb{K}$ and all $A, B \in W$, $a \cdot A + b \cdot B \in W$.*

If we had instead chosen *right* scalar multiplication in $M_{n,m}(\mathbb{K})$, defined as $(A \cdot a)_{ij} = A_{ij} \cdot a$, then $M_{n,m}(\mathbb{K})$ would have become a *right vector space* over $\mathbb{K}$. In a right vector space, scalar multiplication is denoted $a, A \mapsto A \cdot a$. Properties (2) through (4) of Definition 1.7 must be re-written to reflect this notational change. Property (1) is special because the change is more than just notational:

(1') $(A \cdot b) \cdot a = A \cdot (b \cdot a).$

Do you see the difference? The net effect of multiplying $A$ by $b$ and then by $a$ is to multiply $A$ by $ab$ in a left vector space, or by $ba$ in a right vector space.

When $\mathbb{K}$ is a field, the difference between a left and a right vector space over $\mathbb{K}$ is an irrelevant notational distinction, so one speaks simply of "vector spaces". But when $\mathbb{K} = \mathbb{H}$, it makes an essential difference that we are henceforth adopting the convention of left scalar multiplication, and thereby choosing to regard $M_{n,m}(\mathbb{H})$ as a left vector space over $\mathbb{H}$.

## 5. Matrices as linear transformations

One cornerstone of a linear algebra course is the discovery that matrices correspond to linear transformations, and vice versa. We now review that discovery. Extra care is needed when $\mathbb{K} = \mathbb{H}$.

**Definition 1.9.** *Suppose that $V_1$ and $V_2$ are left vector spaces over $\mathbb{K}$. A function $f : V_1 \to V_2$ is called $\mathbb{K}$-**linear** (or simply **linear**) if for all $a, b \in \mathbb{K}$ and all $X, Y \in V_1$,*

$$f(a \cdot X + b \cdot Y) = a \cdot f(X) + b \cdot f(Y).$$

It is natural to identify $\mathbb{K}^n = \{(q_1, ..., q_n) \mid q_i \in \mathbb{K}\}$ with $M_{1,n}(\mathbb{K})$ (horizontal single-row matrices) and thereby regard $\mathbb{K}^n$ as a left vector space over $\mathbb{K}$. Using this identification, there are two potential ways in which matrices might correspond to linear transformations from $\mathbb{K}^n$ to $\mathbb{K}^n$:

**Definition 1.10.** *If $A \in M_n(\mathbb{K})$, define $R_A : \mathbb{K}^n \to \mathbb{K}^n$ and define $L_A : \mathbb{K}^n \to \mathbb{K}^n$ such that for $X \in \mathbb{K}^n$,*

$$R_A(X) = X \cdot A \quad \textit{and} \quad L_A(X) = (A \cdot X^T)^T.$$

For example, if $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \in M_2(\mathbb{R})$, then for $(x, y) \in \mathbb{R}^2$,

$$R_A(x,y) = \begin{pmatrix} x & y \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = (x + 3y, 2x + 4y), \text{ and}$$

$$L_A(x,y) = \left( \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \right)^T = \begin{pmatrix} x + 2y \\ 3x + 4y \end{pmatrix}^T = (x + 2y, 3x + 4y).$$

We first prove that *right* multiplication determines a one-to-one correspondence between linear functions from $\mathbb{K}^n$ to $\mathbb{K}^n$ and matrices.

**Proposition 1.11.**

(1) *For any* $A \in M_n(\mathbb{K})$, $R_A : \mathbb{K}^n \to \mathbb{K}^n$ *is* $\mathbb{K}$*-linear.*

(2) *Each* $\mathbb{K}$*-linear function from* $\mathbb{K}^n$ *to* $\mathbb{K}^n$ *equals* $R_A$ *for some* $A \in M_n(\mathbb{K})$.

**Proof.** To prove (1), notice that for all $a, b \in \mathbb{K}$ and $X, Y \in \mathbb{K}^n$,

$$\begin{aligned} R_A(aX + bY) = (aX + bY) \cdot A &= a(X \cdot A) + b(Y \cdot A) \\ &= a \cdot R_A(X) + b \cdot R_A(Y). \end{aligned}$$

To prove (2), assume that $f : \mathbb{K}^n \to \mathbb{K}^n$ is $\mathbb{K}$-linear. Let $A \in M_n(\mathbb{K})$ denote the matrix whose $i^{\text{th}}$ row is $f(e_i)$, where

$$e_1 = (1, 0, ..., 0), e_2 = (0, 1, 0, ..., 0), ..., e_n = (0, ..., 0, 1)$$

denotes the standard basis for $\mathbb{K}^n$. It's easy to see that $f(e_i) = R_A(e_i)$ for all $i = 1, ..., n$. Since $f$ and $R_A$ are both linear maps and they agree on a basis, we conclude that $f = R_A$. □

We see from the proof that the rows of $A \in M_n(\mathbb{K})$ are the images under $R_A$ of $\{e_1, ..., e_n\}$. Similarly, the columns are the images under $L_A$.

Most linear algebra textbooks use the convention of identifying a matrix $A \in M_n(\mathbb{K})$ with the function $L_A : \mathbb{K}^n \to \mathbb{K}^n$. Unfortunately, this function is necessarily $\mathbb{K}$-linear only when $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

**Proposition 1.12.** *Let* $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

(1) *For any* $A \in M_n(\mathbb{K})$, $L_A : \mathbb{K}^n \to \mathbb{K}^n$ *is* $\mathbb{K}$*-linear.*

(2) *Each* $\mathbb{K}$*-linear function from* $\mathbb{K}^n$ *to* $\mathbb{K}^n$ *equals* $L_A$ *for some* $A \in M_n(\mathbb{K})$.

Proposition 1.12 is an immediate corollary of Proposition 1.11 plus the following easily verified fact:

$$L_A = R_{A^T} \text{ for all } A \in M_n(\mathbb{R}) \text{ or } A \in M_n(\mathbb{C}).$$

Our previous decision to consider $\mathbb{H}^n$ as a *left* vector space over $\mathbb{H}$ forces us now to use the correspondence $A \leftrightarrow R_A$ between matrices and linear transformations (rather than $A \leftrightarrow L_A$), at least when we wish to include $\mathbb{K} = \mathbb{H}$ in our discussion.

Under either correspondence between matrices and transformations, matrix multiplication corresponds to composition of transformations, since:

$$L_A(L_B(X)) = L_{A\cdot B}(X) \text{ and } R_A(R_B(X)) = R_{B\cdot A}(X).$$

In a linear algebra course, this is one's first indication that the initially unmotivated definition of matrix multiplication is in fact quite natural.

## 6. The general linear groups

The set $M_n(\mathbb{K})$ is not a group under matrix multiplication because some matrices do not have multiplicative inverses. For example, if $A \in M_n(\mathbb{K})$ has all entries zero, then $A$ has no multiplicative inverse; that is, there is no matrix $B$ for which $AB = BA = I$. However, the elements of $M_n(\mathbb{K})$ which do have inverses form a very important group whose subgroups are the main topic of this text.

**Definition 1.13.** *The **general linear group** over $\mathbb{K}$ is:*

$$GL_n(\mathbb{K}) = \{A \in M_n(\mathbb{K}) \mid \exists B \in M_n(\mathbb{K}) \text{ such that } AB = BA = I\}.$$

Such a matrix $B$ is the multiplicative inverse of $A$ and is therefore denoted $A^{-1}$. As its name suggests, $GL_n(\mathbb{K})$ is a group under the operation of matrix multiplication (why?). The following more visual characterization of the general linear group is often useful:

**Proposition 1.14.**

$$GL_n(\mathbb{K}) = \{A \in M_n(\mathbb{K}) \mid R_A : \mathbb{K}^n \to \mathbb{K}^n \text{ is a linear isomorphism}\}.$$

For $A \in M_n(\mathbb{K})$, $R_A$ is always linear; it is called an isomorphism if it is invertible (or equivalently, surjective, or equivalently, injective). Thus, general linear matrices correspond to motions of $\mathbb{K}^n$ with no collapsing.

**Proof.** If $A \in GL_n(\mathbb{K})$ and $B$ is such that $BA = I$, then

$$R_A \circ R_B = R_{BA} = R_I = \text{id (the identity)},$$

so $R_A$ has inverse $R_B$.

Conversely, let $A \in M_n(\mathbb{K})$ be such that $R_A$ is invertible. The map $(R_A)^{-1}$ is linear, which can be seen by applying $R_A$ to both sides of the following equation:

$$(R_A)^{-1}(aX + bY) \stackrel{?}{=} a(R_A)^{-1}(X) + b(R_A)^{-1}(Y).$$

Since every linear map is represented by a matrix, $(R_A)^{-1} = R_B$ for some $B \in M_n(\mathbb{K})$. Therefore, $R_{BA} = R_A \circ R_B = \text{id}$, which implies $BA = I$. Similarly, $R_{AB} = R_B \circ R_A = \text{id}$, which implies $AB = I$. □

The following well-known fact from linear algebra provides yet another useful description of the general linear group, at least when $\mathbb{K} \neq \mathbb{H}$:

**Proposition 1.15.** *If* $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$*, then*

$$GL_n(\mathbb{K}) = \{A \in M_n(\mathbb{K}) \mid \det(A) \neq 0\}.$$

In fact, the elements of the inverse of a matrix can be described explicitly in terms of the determinant of the matrix and its minors:

**Proposition 1.16** (Cramer's rule)**.** *Let* $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$*. Using the notation of Equation* 1.5,

$$(A^{-1})_{ij} = (-1)^{i+j} \frac{\det(A[j,i])}{\det(A)}.$$

## 7. Change of basis via conjugation

In this section, we review a basic fact from linear algebra: a conjugate of a matrix represents the same linear transformation as the matrix, but in a different basis.

Let $\mathfrak{g}$ denote an $n$-dimensional (left) vector space over $\mathbb{K}$. Then $\mathfrak{g}$ is isomorphic to $\mathbb{K}^n$. In fact, there are many isomorphisms from $\mathfrak{g}$ to $\mathbb{K}^n$. For any ordered basis $V = \{v_1, ..., v_n\}$ of $\mathfrak{g}$, the following is an isomorphism:

$$(c_1v_1 + \cdots + c_nv_n) \mapsto (c_1, ..., c_n). \tag{1.7}$$

Every isomorphism from $\mathfrak{g}$ to $\mathbb{K}^n$ has this form for some ordered basis of $\mathfrak{g}$, so choosing an isomorphism amounts to choosing an ordered basis. In practice, there is typically no choice of basis that is more natural than the other choices. To convince yourself of this, consider the case where $\mathfrak{g}$ is an arbitrary subspace of $\mathbb{K}^m$ for some $m > n$.

Now suppose that $f : \mathfrak{g} \to \mathfrak{g}$ is a linear transformation. In order to identify $f$ with a matrix, we must first choose an ordered basis $V$ of $\mathfrak{g}$. We use this basis to identify $\mathfrak{g} \cong \mathbb{K}^n$ and thereby to regard $f$ as a linear transformation from $\mathbb{K}^n$ to $\mathbb{K}^n$, which can be represented as $R_A$ for some $A \in M_n(\mathbb{K})$. A crucial point is that $A$ depends on the choice of ordered basis. To emphasize this dependence, we say that "$A$ represents $f$ in the basis $V$ (via right-multiplication)." We would like to determine which matrix represents $f$ in a different basis.

To avoid cumbersome notation, we will simplify this problem without really losing generality. Suppose that $f : \mathbb{K}^n \to \mathbb{K}^n$ is a linear transformation. We know that $f = R_A$ for some $A \in M_n(\mathbb{K})$. Translating this sentence into our new terminology, we say that "$A$ represents $f$ in the standard basis of $\mathbb{K}^n$," which is:

$$\{e_1 = (1, 0, ..., 0), e_2 = (0, 1, 0, ..., 0), ..., e_n = (0, ..., 0, 1)\}.$$

Now let $V = \{v_1, ..., v_n\}$ denote an arbitrary basis of $\mathbb{K}^n$. We seek the matrix that represents $f$ in the basis $V$. First, we let $g \in GL_n(\mathbb{K})$ denote the matrix whose rows are $v_1, v_2, ..., v_n$. We call $g$ the **change of basis matrix**. To understand why, notice that $e_i g = v_i$ for each $i = 1, ..., n$. So,

$$(c_1, ..., c_n) \cdot g = (c_1 e_1 + \cdots + c_n e_n) \cdot g = c_1 v_1 + \cdots + c_n v_n.$$

By Equation 1.7, the vector $c_1 v_1 + \cdots + c_n v_n \in \mathbb{K}^n$ is represented in the basis $V$ as $(c_1, ..., c_n)$. Thus, $R_g : \mathbb{K}^n \to \mathbb{K}^n$ translates between $V$ and the standard basis. For $X \in \mathbb{K}^n$, $R_g(X)$ represents in the standard basis the same vector that $X$ represents in $V$. Further, $R_{g^{-1}}(X)$ represents in $V$ the same vector that $X$ represents in the standard basis.

**Proposition 1.17.** *$gAg^{-1}$ represents $f$ in the basis $V$.*

**Proof.** Let $X = (c_1, ..., c_n)$, which represents $c_1 v_1 + \cdots + c_n v_n$ in $V$. We must show that $R_{gAg^{-1}}(X)$ represents $(c_1 v_1 + \cdots + c_n v_n) \cdot A$ in

$V$. This follows from:

$$\begin{aligned} R_{gAg^{-1}}(X) = (c_1, ..., c_n)gAg^{-1} &= (c_1v_1 + \cdots + c_nv_n)Ag^{-1} \\ &= R_{g^{-1}}((c_1v_1 + \cdots + c_nv_n) \cdot A). \end{aligned}$$

□

Proposition 1.17 can be summarized in the following way: for any $A \in M_n(\mathbb{K})$ and any $g \in GL_n(\mathbb{K})$, the matrix $gAg^{-1}$ represents $R_A$ in the basis $\{e_1g, ..., e_ng\}$.

The basic idea of the proof was simple enough: the transformation $R_{gAg^{-1}} = R_{g^{-1}} \circ R_A \circ R_g$ first translates into the standard basis, then performs the transformation associated to $A$, then translates back.

This key result requires only slight modification when representing linear transformations using *left* matrix multiplication when $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$: for any $A \in M_n(\mathbb{K})$ and any $g \in GL_n(\mathbb{K})$, the matrix $g^{-1}Ag$ represents $L_A$ in the basis $\{ge_1, ..., ge_n\}$ (via left multiplication). The proof idea is the same: $L_{g^{-1}Ag} = L_{g^{-1}} \circ L_A \circ L_g$ first translates into the standard basis, then performs the transformation associated to $A$, then translates back.

## 8. Exercises

**Ex. 1.1.** Describe a natural 1-to-1 correspondence between elements of $SO(3)$ and elements of

$$T^1S^2 = \{(p, v) \in \mathbb{R}^3 \times \mathbb{R}^3 \mid |p| = |v| = 1 \text{ and } p \perp v\},$$

which can be thought of as the collection of all unit-length vectors $v$ tangent to all points $p$ of $S^2$. Compare to Question 1.2.

**Ex. 1.2.** Prove Equation 1.3.

**Ex. 1.3.** Prove Equation 1.4.

**Ex. 1.4.** Let $A, B \in M_n(\mathbb{K})$. Prove that if $AB = I$, then $BA = I$.

**Ex. 1.5.** Suppose that the determinant of $A \in M_n(\mathbb{H})$ were defined as in Equation 1.5. Show for $A = \begin{pmatrix} \mathbf{i} & \mathbf{j} \\ \mathbf{i} & \mathbf{j} \end{pmatrix} \in M_2(\mathbb{H})$ that $\det(A) \neq 0$ but $R_A : \mathbb{H}^2 \to \mathbb{H}^2$ is not invertible.

**Ex. 1.6.** Find $B \in M_2(\mathbb{R})$ such that $R_B : \mathbb{R}^2 \to \mathbb{R}^2$ is a counter-clockwise rotation through an angle $\theta$.

**Ex. 1.7.** Describe all elements $A \in GL_n(\mathbb{R})$ with the property that $AB = BA$ for all $B \in GL_n(\mathbb{R})$.

**Ex. 1.8.** Let $SL_2(\mathbb{Z})$ denote the set of all 2 by 2 matrices with integer entries and with determinant 1. Prove that $SL_2(\mathbb{Z})$ is a subgroup of $GL_2(\mathbb{R})$. Is $SL_n(\mathbb{Z})$ (defined analogously) a subgroup of $GL_n(\mathbb{R})$?

**Ex. 1.9.** Describe the product of two matrices in $M_6(\mathbb{K})$ which both have the form:

$$\begin{pmatrix} a & b & 0 & 0 & 0 & 0 \\ c & d & 0 & 0 & 0 & 0 \\ 0 & 0 & e & f & g & 0 \\ 0 & 0 & h & i & j & 0 \\ 0 & 0 & k & l & m & 0 \\ 0 & 0 & 0 & 0 & 0 & n \end{pmatrix}$$

Describe a general rule for the product of two matrices with the same **block form**.

**Ex. 1.10.** If $G_1 \subset GL_{n_1}(\mathbb{K})$ and $G_2 \subset GL_{n_2}(\mathbb{K})$ are subgroups, describe a subgroup of $GL_{n_1+n_2}(\mathbb{K})$ that is isomorphic to $G_1 \times G_2$.

**Ex. 1.11.** Show by example that for $A \in M_n(\mathbb{H})$, $L_A : \mathbb{H}^n \to \mathbb{H}^n$ is not necessarily $\mathbb{H}$-linear.

**Ex. 1.12.** Define the *real* and *imaginary* parts of a quaternion as follows:

$$\mathrm{Re}(a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) = a$$

$$\mathrm{Im}(a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) = b\mathbf{i} + c\mathbf{j} + d\mathbf{k}.$$

Let $q_1 = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}$ and $q_2 = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}$ be *purely imaginary* quaternions in $\mathbb{H}$. Prove that $-\mathrm{Re}(q_1 \cdot q_2)$ is their vector dot product in $\mathbb{R}^3 = \mathrm{span}\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ and $\mathrm{Im}(q_1 \cdot q_2)$ is their vector cross product.

**Ex. 1.13.** Prove that non-real elements $q_1, q_2 \in \mathbb{H}$ commute if and only if their imaginary parts are parallel; that is, $\mathrm{Im}(q_1) = \lambda \cdot \mathrm{Im}(q_2)$ for some $\lambda \in \mathbb{R}$.

**Ex. 1.14.** Characterize the pairs $q_1, q_2 \in \mathbb{H}$ which anti-commute, meaning that $q_1 q_2 = -q_2 q_1$.

**Ex. 1.15.** If $q \in \mathbb{H}$ satisfies $q\mathbf{i} = \mathbf{i}q$, prove that $q \in \mathbb{C}$.

**Ex. 1.16.** Prove that complex multiplication in $\mathbb{C} \cong \mathbb{R}^2$ does not extend to a multiplication operation on $\mathbb{R}^3$ that makes $\mathbb{R}^3$ into a real division algebra.

**Ex. 1.17.** Describe a subgroup of $GL_{n+1}(\mathbb{R})$ that is isomorphic to the group $\mathbb{R}^n$ under the operation of vector-addition.

**Ex. 1.18.** If $\lambda \in \mathbb{H}$ commutes with every element of $\mathbb{H}$, prove that $\lambda \in \mathbb{R}$.

Chapter 2

# All matrix groups are real matrix groups

This book is about subgroups of the general linear groups. In this chapter, we prove that every subgroup of $GL_n(\mathbb{C})$ or $GL_n(\mathbb{H})$ is isomorphic to a subgroup of $GL_m(\mathbb{R})$ for some $m$. Thus, this book might as well be about subgroups of the *real* general linear group. The result is an immediate consequence of:

**Theorem 2.1.**

(1) $GL_n(\mathbb{C})$ *is isomorphic to a subgroup of* $GL_{2n}(\mathbb{R})$.

(2) $GL_n(\mathbb{H})$ *is isomorphic to a subgroup of* $GL_{2n}(\mathbb{C})$.

It follows that $GL_n(\mathbb{H})$ is isomorphic to a subgroup of $GL_{4n}(\mathbb{R})$. We will prove Theorem 2.1 by constructing injective homomorphisms:

$$\rho_n : GL_n(\mathbb{C}) \to GL_{2n}(\mathbb{R}) \quad \text{and} \quad \Psi_n : GL_n(\mathbb{H}) \to GL_{2n}(\mathbb{C}).$$

These homomorphisms play an important role in the remainder of the text.

Many important groups are much more naturally regarded as subgroups of $GL_n(\mathbb{H})$ or $GL_n(\mathbb{C})$ rather than of $GL_m(\mathbb{R})$, so the theorem does not obviate our future need to consider the cases $\mathbb{K} = \mathbb{C}$ and $\mathbb{K} = \mathbb{H}$.

## 1. Complex matrices as real matrices

In Exercise 1.6, you showed for the matrix

$$B = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \in M_2(\mathbb{R})$$

that $R_B : \mathbb{R}^2 \to \mathbb{R}^2$ is a counterclockwise rotation through angle $\theta$. In fact, standard trigonometric identities give that for all $r, \phi \in \mathbb{R}$:

$$R_B(r\cos\phi, r\sin\phi) = (r\cos(\theta+\phi), r\sin(\theta+\phi)).$$

Compare this to the matrix $A = (e^{\mathbf{i}\theta}) \in M_1(\mathbb{C})$. For this matrix, $R_A : \mathbb{C}^1 \to \mathbb{C}^1$ is also a counterclockwise rotation through angle $\theta$, since

$$R_A(re^{\mathbf{i}\phi}) = re^{\mathbf{i}(\theta+\phi)}.$$

Thus, $A \in M_1(\mathbb{C})$ and $B \in M_2(\mathbb{R})$ "represent the same motion".

More generally, we wish to construct a function

$$\rho_n : M_n(\mathbb{C}) \to M_{2n}(\mathbb{R})$$

that sends $A \in M_n(\mathbb{C})$ to the matrix $B \in M_{2n}(\mathbb{R})$ that "represents the same motion". More precisely, every $A \in M_n(\mathbb{C})$ corresponds to a linear transformation $R_A : \mathbb{C}^n \to \mathbb{C}^n$. This transformation can instead be thought of as a transformation from $\mathbb{R}^{2n}$ to $\mathbb{R}^{2n}$, since $\mathbb{R}^{2n}$ is naturally identified with $\mathbb{C}^n$ via the bijection $f_n : \mathbb{C}^n \to \mathbb{R}^{2n}$ defined as:

$$f_n(a_1 + b_1\mathbf{i}, a_2 + b_2\mathbf{i}, ..., a_n + b_n\mathbf{i}) = (a_1, b_1, a_2, b_2, ..., a_n, b_n).$$

This transformation from $\mathbb{R}^{2n}$ to $\mathbb{R}^{2n}$ is represented as $R_B$ for some $B \in M_{2n}(\mathbb{R})$.

How do we determine $B$ from $A$? Asked differently, how do we define a function

$$\rho_n : M_n(\mathbb{C}) \to M_{2n}(\mathbb{R})$$

such that the following diagram commutes for all $A \in M_n(\mathbb{C})$:

$$\begin{tikzcd} \mathbb{C}^n \arrow[r, "f_n"] \arrow[d, "R_A"'] & \mathbb{R}^{2n} \arrow[d, "R_{\rho_n(A)}"] \\ \mathbb{C}^n \arrow[r, "f_n"] & \mathbb{R}^{2n} \end{tikzcd} \tag{2.1}$$

(the diagram is said to **commute** if $R_{\rho_n(A)} \circ f_n = f_n \circ R_A$; that is, if right-then-down equals down-then-right). When $n = 1$, it is straightforward to check that the function $\rho_1 : M_1(\mathbb{C}) \to M_2(\mathbb{R})$ defined as follows makes diagram 2.1 commute:

$$\rho_1(a + b\mathbf{i}) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Notice that $\rho_1$ relates the matrices $A$ and $B$ of the previous discussion, since

$$\rho_1(e^{i\theta}) = \rho_1(\cos\theta + \mathbf{i}\sin\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \in M_2(\mathbb{R}).$$

For $A \in M_n(\mathbb{C})$ with $n > 1$, we build $\rho_n(A)$ out of 2-by-2 blocks equal to $\rho_1$ applied to the entries of $A$. For example,

$$\rho_2 \begin{pmatrix} a + b\mathbf{i} & c + d\mathbf{i} \\ e + f\mathbf{i} & h + j\mathbf{i} \end{pmatrix} = \begin{pmatrix} a & b & c & d \\ -b & a & -d & c \\ e & f & h & j \\ -f & e & -j & h \end{pmatrix},$$

and so on. In Exercise 2.1, you will prove that this definition of $\rho_n$ makes diagram 2.1 commute.

**Proposition 2.2.** *For all $\lambda \in \mathbb{R}$ and $A, B \in M_n(\mathbb{C})$,*

(1) $\rho_n(\lambda \cdot A) = \lambda \cdot \rho_n(A)$.
(2) $\rho_n(A + B) = \rho_n(A) + \rho_n(B)$.
(3) $\rho_n(A \cdot B) = \rho_n(A) \cdot \rho_n(B)$.

**Proof.** Parts (1) and (2) are immediate from definition. For part (3), consider the commutative diagram:

$$\begin{tikzcd}
\mathbb{C}^n \arrow[r, "f_n"] \arrow[d, "R_A"'] & \mathbb{R}^{2n} \arrow[d, "R_{\rho_n(A)}"] \\
\mathbb{C}^n \arrow[r, "f_n"] \arrow[d, "R_B"'] & \mathbb{R}^{2n} \arrow[d, "R_{\rho_n(B)}"] \\
\mathbb{C}^n \arrow[r, "f_n"] & \mathbb{R}^{2n}
\end{tikzcd}$$

The composition of the two down-arrows on the right is

$$R_{\rho_n(B)} \circ R_{\rho_n(A)} = R_{\rho_n(A)\cdot\rho_n(B)}.$$

On the other hand, since on the left $R_B \circ R_A = R_{AB}$, this composition on the right also equals $R_{\rho_n(AB)}$. In summary,

$$R_{\rho_n(A)\cdot\rho_n(B)} = R_{\rho_n(AB)},$$

which implies that $\rho_n(A) \cdot \rho_n(B) = \rho_n(AB)$. □

It is easy to see that $\rho_n : M_n(\mathbb{C}) \to M_{2n}(\mathbb{R})$ is injective but not surjective.

**Definition 2.3.** *Matrices of $M_{2n}(\mathbb{R})$ in the image of $\rho_n$ are called **complex-linear** real matrices.*

The terminology is justified by the following proposition, whose proof is immediate.

**Proposition 2.4.** *$B \in M_{2n}(\mathbb{R})$ is complex-linear if and only if the composition $f_n^{-1} \circ R_B \circ f_n : \mathbb{C}^n \to \mathbb{C}^n$ is a $\mathbb{C}$-linear transformation.*

This composition $F = f_n^{-1} \circ R_B \circ f_n$ is depicted as:

$$\begin{array}{ccc} \mathbb{C}^n & \xrightarrow{f_n} & \mathbb{R}^{2n} \\ & & \downarrow R_B \\ \mathbb{C}^n & \xleftarrow{f_n^{-1}} & \mathbb{R}^{2n} \end{array}$$

Notice that $F$ is always $\mathbb{R}$-linear (which makes sense because $\mathbb{C}^n$ can be regarded as a vector space over $\mathbb{R}$). It is $\mathbb{C}$-linear if and only if $F(\mathbf{i} \cdot X) = \mathbf{i} \cdot F(X)$ for all $X \in \mathbb{C}^n$. So the complex linear real matrices are the ones that "commute with $\mathbf{i}$" in this sense. There is an important way to re-describe this idea of a real matrix commuting with $\mathbf{i}$. Define $J_{2n} = \rho_n(\mathbf{i} \cdot I)$, so for example,

$$J_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}.$$

Notice that $J_{2n}^2 = -1 \cdot I$ and that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{C}^n & \xrightarrow{f_n} & \mathbb{R}^{2n} \\ {\scriptstyle R_{\mathbf{i}\cdot I} = (\text{scalar mult. by } \mathbf{i})}\downarrow & & \downarrow {\scriptstyle R_{J_{2n}}} \\ \mathbb{C}^n & \xrightarrow{f_n} & \mathbb{R}^{2n} \end{array}$$

The matrix $J_{2n}$ is called the **standard complex structure** on $\mathbb{R}^{2n}$. Why? Because, compared to $\mathbb{R}^{2n}$, the space $\mathbb{C}^n$ has the additional structure of scalar-multiplication by $\mathbf{i}$. This extra structure is mimicked in $\mathbb{R}^{2n}$ by $R_{J_{2n}}$. This allows an improved verbalization of the above-indicated idea that complex-linear real matrices "commute with $\mathbf{i}$":

**Proposition 2.5.** *$B \in M_{2n}(\mathbb{R})$ is complex-linear if and only if*

$$B \cdot J_{2n} = J_{2n} \cdot B.$$

**Proof.** Suppose that $B \in M_{2n}(\mathbb{R})$ is complex-linear, which means there is a matrix $A \in M_n(\mathbb{C})$ such that $\rho_n(A) = B$. The following diagram commutes:

$$\begin{array}{ccc} \mathbb{C}^n & \xrightarrow{f_n} & \mathbb{R}^{2n} \\ {\scriptstyle (\text{scalar mult. by } \mathbf{i})}\downarrow & & \downarrow {\scriptstyle R_{J_{2n}}} \\ \mathbb{C}^n & \xrightarrow{f_n} & \mathbb{R}^{2n} \\ {\scriptstyle R_A}\downarrow & & \downarrow {\scriptstyle R_B} \\ \mathbb{C}^n & \xrightarrow{f_n} & \mathbb{R}^{2n} \\ {\scriptstyle (\text{scalar mult. by } \mathbf{i})}\downarrow & & \downarrow {\scriptstyle R_{J_{2n}}} \\ \mathbb{C}^n & \xrightarrow{f_n} & \mathbb{R}^{2n} \end{array}$$

The composition of the three downward arrows on the left equals $R_{iAi} = R_{-A}$, so the composition of the three downward arrows on the right must equal $R_{\rho_n(-A)} = R_{-B}$. Therefore:

$$R_{-B} = R_{J_{2n}BJ_{2n}}.$$

It follows that $-B = J_{2n}BJ_{2n}$. Since $J_{2n}^2 = -I$, this implies that $B \cdot J_{2n} = J_{2n} \cdot B$.

The other direction is similar and is left to the reader. $\square$

## 2. Quaternionic matrices as complex matrices

The results in this section are analogous to results from the previous section, so we discuss them only briefly. The main idea is to think of elements of $M_n(\mathbb{H})$ as transformations of $\mathbb{C}^{2n}$ or $\mathbb{R}^{4n}$.

Recall from Section 3 of Chapter 1 that any element of $\mathbb{H}$ can be expressed as $z + w\mathbf{j}$ for some $z, w \in \mathbb{C}$. Using this, there is a natural bijection $g_n : \mathbb{H}^n \to \mathbb{C}^{2n}$ defined as

$$g_n(z_1 + w_1\mathbf{j}, z_2 + w_2\mathbf{j}, ..., z_n + w_n\mathbf{j}) = (z_1, w_1, z_2, w_2, ..., z_n, w_n).$$

Our goal is to define an injective map

$$\Psi_n : M_n(\mathbb{H}) \to M_{2n}(\mathbb{C})$$

such that the following diagram commutes for all $A \in M_n(\mathbb{H})$:

$$\begin{tikzcd} \mathbb{H}^n \arrow[r, "g_n"] \arrow[d, "R_A"'] & \mathbb{C}^{2n} \arrow[d, "R_{\Psi_n(A)}"] \\ \mathbb{H}^n \arrow[r, "g_n"] & \mathbb{C}^{2n} \end{tikzcd} \tag{2.2}$$

The solution when $n = 1$ is:

$$\Psi_1(z + w\mathbf{j}) = \begin{pmatrix} z & w \\ -\overline{w} & \overline{z} \end{pmatrix},$$

where overlines denote complex conjugation, defined as $\overline{a + b\mathbf{i}} = a - b\mathbf{i}$. An alternative way to express $\Psi_1$ is:

$$\Psi_1(a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) = \begin{pmatrix} a + b\mathbf{i} & c + d\mathbf{i} \\ -c + d\mathbf{i} & a - b\mathbf{i} \end{pmatrix}.$$

For $n > 1$, define $\Psi_n$ in terms of $\Psi_1$ exactly the way $\rho_n$ was defined in terms of $\rho_1$. For example,

$$\Psi_2 \begin{pmatrix} a_{11} + b_{11}\mathbf{i} + c_{11}\mathbf{j} + d_{11}\mathbf{k} & a_{12} + b_{12}\mathbf{i} + c_{12}\mathbf{j} + d_{12}\mathbf{k} \\ a_{21} + b_{21}\mathbf{i} + c_{21}\mathbf{j} + d_{21}\mathbf{k} & a_{22} + b_{22}\mathbf{i} + c_{22}\mathbf{j} + d_{22}\mathbf{k} \end{pmatrix}$$
$$= \begin{pmatrix} a_{11} + b_{11}\mathbf{i} & c_{11} + d_{11}\mathbf{i} & a_{12} + b_{12}\mathbf{i} & c_{12} + d_{12}\mathbf{i} \\ -c_{11} + d_{11}\mathbf{i} & a_{11} - b_{11}\mathbf{i} & -c_{12} + d_{12}\mathbf{i} & a_{12} - b_{12}\mathbf{i} \\ a_{21} + b_{21}\mathbf{i} & c_{21} + d_{21}\mathbf{i} & a_{22} + b_{22}\mathbf{i} & c_{22} + d_{22}\mathbf{i} \\ -c_{21} + d_{21}\mathbf{i} & a_{21} - b_{21}\mathbf{i} & -c_{22} + d_{22}\mathbf{i} & a_{22} - b_{22}\mathbf{i} \end{pmatrix}.$$

Matrices of $M_{2n}(\mathbb{C})$ in the image of $\Psi_n$ are called **quaternionic-linear** complex matrices.

**Proposition 2.6.** *For all $\lambda \in \mathbb{R}$ and $A, B \in M_n(\mathbb{H})$,*

(1) $\Psi_n(\lambda \cdot A) = \lambda \cdot \Psi_n(A)$.
(2) $\Psi_n(A + B) = \Psi_n(A) + \Psi_n(B)$.
(3) $\Psi_n(A \cdot B) = \Psi_n(A) \cdot \Psi_n(B)$.

*Further, $B \in M_{2n}(\mathbb{C})$ is quaternionic-linear if and only if the composition $F = g_n^{-1} \circ R_B \circ g_n : \mathbb{H}^n \to \mathbb{H}^n$ is an $\mathbb{H}$-linear transformation.*

The composition $F$ is depicted in the following diagram:

$$\begin{tikzcd}
\mathbb{H}^n \arrow[r, "g_n"] & \mathbb{C}^{2n} \arrow[d, "R_B"] \\
\mathbb{H}^n & \mathbb{C}^{2n} \arrow[l, "g_n^{-1}"']
\end{tikzcd}$$

Putting it together, we have injective maps $\rho_n : M_n(\mathbb{C}) \to M_{2n}(\mathbb{R})$ and $\Psi_n : M_n(\mathbb{H}) \to M_{2n}(\mathbb{C})$ such that the following commutes for all $A \in M_n(\mathbb{H})$:

$$\begin{tikzcd}
\mathbb{H}^n \arrow[r, "g_n"] \arrow[d, "R_A"] & \mathbb{C}^{2n} \arrow[r, "f_{2n}"] \arrow[d, "R_{\Psi_n(A)}"] & \mathbb{R}^{4n} \arrow[d, "R_{(\rho_{2n} \circ \Psi_n)(A)}"] \\
\mathbb{H}^n \arrow[r, "g_n"] & \mathbb{C}^{2n} \arrow[r, "f_{2n}"] & \mathbb{R}^{4n}
\end{tikzcd}$$

Matrices of $M_{4n}(\mathbb{R})$ in the image of $(\rho_{2n} \circ \Psi_n) : M_n(\mathbb{H}) \to M_{4n}(\mathbb{R})$ are called **quaternionic-linear** real matrices.

**Proposition 2.7.** *The following are equivalent for $B \in M_{4n}(\mathbb{R})$.*

(1) *$B$ is quaternionic-linear.*
(2) *$B$ commutes with both $\mathcal{I}_{4n}$ and $\mathcal{J}_{4n}$.*
(3) *$(g_n^{-1} \circ f_{2n}^{-1} \circ R_B \circ f_{2n} \circ g_n) : \mathbb{H}^n \to \mathbb{H}^n$ is $\mathbb{H}$-linear.*

The composition in part (3) is depicted as:

$$\begin{tikzcd}
\mathbb{H}^n \arrow[r, "g_n"] & \mathbb{C}^{2n} \arrow[r, "f_{2n}"] & \mathbb{R}^{4n} \arrow[d, "R_B"] \\
\mathbb{H}^n & \mathbb{C}^{2n} \arrow[l, "g_n^{-1}"'] & \mathbb{R}^{4n} \arrow[l, "f_{2n}^{-1}"']
\end{tikzcd}$$

The matrices $\mathcal{I}_{4n}$ and $\mathcal{J}_{4n}$ from the proposition are defined such that these diagrams commute:

$$\begin{array}{ccc} \mathbb{H}^n & \xrightarrow{(f_{2n}\circ g_n)} & \mathbb{R}^{4n} \\ \Big\downarrow \text{(scalar mult. } \mathbf{i}) & & \Big\downarrow R_{\mathcal{I}_{4n}} \\ \mathbb{H}^n & \xrightarrow{(f_{2n}\circ g_n)} & \mathbb{R}^{4n} \end{array} \qquad \begin{array}{ccc} \mathbb{H}^n & \xrightarrow{(f_{2n}\circ g_n)} & \mathbb{R}^{4n} \\ \Big\downarrow \text{(scalar mult. } \mathbf{j}) & & \Big\downarrow R_{\mathcal{J}_{4n}} \\ \mathbb{H}^n & \xrightarrow{(f_{2n}\circ g_n)} & \mathbb{R}^{4n} \end{array}$$

"Scalar mult. **i**" means *left* scalar multiplication by **i**, and similarly for **j**. The analogy with Section 1 is imperfect, since $\mathcal{I}_{4n}$ and $\mathcal{J}_{4n}$ do not equal $(\rho_{2n} \circ \Psi_n)(\mathbf{i}I)$ and $(\rho_{2n} \circ \Psi_n)(\mathbf{j}I)$ (why?). The correct choice for $\mathcal{I}_4$ and $\mathcal{J}_4$ turns out to be:

$$\mathcal{I}_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \qquad \mathcal{J}_4 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

The correct choice for $\mathcal{I}_{4n}$ (respectively $\mathcal{J}_{4n}$) has block-form with blocks of $\mathcal{I}_4$ (respectively $\mathcal{J}_4$) along the diagonal.

## 3. Restricting to the general linear groups

**Proposition 2.8.** *The image under $\rho_n$ or $\Psi_n$ of an invertible matrix is an invertible matrix.*

**Proof.** Let $A \in M_n(\mathbb{C})$. Then,

$$\begin{aligned} A \in GL_n(\mathbb{C}) \quad &\iff \quad R_A : \mathbb{C}^n \to \mathbb{C}^n \text{ is bijective} \\ &\iff \quad R_{\rho_n(A)} : \mathbb{R}^{2n} \to \mathbb{R}^{2n} \text{ is bijective} \\ &\iff \quad \rho_n(A) \in GL_{2n}(\mathbb{R}). \end{aligned}$$

The argument for $\Psi_n$ is similar. □

Because of this proposition, we can restrict $\rho_n$ and $\Psi_n$ to maps between the general linear groups:

$$\begin{aligned} \rho_n &: GL_n(\mathbb{C}) \to GL_{2n}(\mathbb{R}), \\ \Psi_n &: GL_n(\mathbb{H}) \to GL_{2n}(\mathbb{C}). \end{aligned}$$

By part (3) of Propositions 2.2 and 2.6, these maps are injective homomorphisms between the general linear groups. Theorem 2.1 is an immediate consequence of the existence of these injective homomorphisms.

Subsequent chapters contain many uses for the homomorphisms $\rho_n$ and $\Psi_n$. As a first application, we now use $\Psi_n$ to define the **determinant of a quaternionic matrix**. It turns out that there is no good way to define a quaternionic-valued determinant function on $M_n(\mathbb{H})$ (as indicated by Exercise 1.5). We will settle for a complex-valued determinant, namely, the composition

$$\det \circ \Psi_n : M_n(\mathbb{H}) \to \mathbb{C}.$$

For $A \in M_n(\mathbb{H})$ we will write $\det(A)$ to mean $\det(\Psi_n(A))$. It is obvious that $\det(I) = 1$ and $\det(A \cdot B) = \det(A) \cdot \det(B)$ for all $A, B \in M_n(\mathbb{H})$. Also, Proposition 1.15 extends to the $\mathbb{K} = \mathbb{H}$ case:

**Proposition 2.9.** $GL_n(\mathbb{H}) = \{A \in M_n(\mathbb{H}) \mid \det(A) \neq 0\}$.

**Proof.** Let $A \in M_n(\mathbb{H})$. As in the proof of Proposition 2.8, we have $A \in GL_n(\mathbb{H})$ if and only if $\Psi_n(A) \in GL_{2n}(\mathbb{C})$, which is equivalent to $\det(\Psi_n(A)) \neq 0$. □

So now for all $K \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$, one can characterize the non-invertible matrices $A \in M_n(\mathbb{K})$ as those which satisfy $\det(A) = 0$, which is a polynomial equation in the entries of $A$.

The determinant of a quaternionic matrix is defined to be a complex number; surprisingly, it is always a real number:

**Proposition 2.10.** *For all* $A \in M_n(\mathbb{H})$, $\det(A) \in \mathbb{R}$.

The proof would take us too far afield from our topic, so we refer the reader to [**8**].

## 4. Exercises

**Ex. 2.1.** Prove that definition of $\rho_n$ in the text makes diagram 2.1 commute.

**Ex. 2.2.** Prove Proposition 2.4.

**Ex. 2.3.** Prove Proposition 2.6.

**Ex. 2.4.** Prove Proposition 2.7.

**Ex. 2.5.** Prove that for any $A \in M_1(\mathbb{H})$, $\det(A) \in \mathbb{R}$.

**Ex. 2.6.** Prove that $SL_n(\mathbb{H}) = \{A \in GL_n(\mathbb{H}) \mid \det(A) = 1\}$ is a subgroup. Describe a natural bijection between elements of $SL_1(\mathbb{H})$ and points of the 3-dimensional sphere $S^3$.

**Ex. 2.7.** Consider the following alternative way to define the function $f_n : \mathbb{C}^n \to \mathbb{R}^{2n}$:

$$f_n(a_1 + b_1\mathbf{i}, ..., a_n + b_n\mathbf{i}) = (a_1, ..., a_n, b_1, ..., b_n).$$

Using this definition, how must $\rho_n$ be defined so that diagram 2.1 commutes? How must $J_{2n}$ be defined so that Proposition 2.5 is true?

**Ex. 2.8.** Is it possible to find a matrix $J \in M_{2n}(\mathbb{C})$ such that the following diagram commutes?

$$\begin{tikzcd} \mathbb{H}^n \arrow[r, "g_n"] \arrow[d, "(\text{scalar mult. by } \mathbf{j})"'] & \mathbb{C}^{2n} \arrow[d, "R_J"] \\ \mathbb{H}^n \arrow[r, "g_n"] & \mathbb{C}^{2n} \end{tikzcd}$$

**Ex. 2.9.** Show that the image $\rho_n(M_n(\mathbb{C})) \subset M_{2n}(\mathbb{R})$ is a real vector subspace. What is its dimension?

**Ex. 2.10.** Are the matrices $\mathcal{I}_{4n}$ and $\mathcal{J}_{4n}$ defined in Proposition 2.7 quaternionic-linear?

**Ex. 2.11.** Is part (1) of Proposition 2.6 true when $\lambda \in \mathbb{C}$?

**Ex. 2.12.** Let $q \in \mathbb{H}$, and define $\mathbb{C} \cdot q = \{\lambda \cdot q \mid \lambda \in \mathbb{C}\} \subset \mathbb{H}$ and $q \cdot \mathbb{C} = \{q \cdot \lambda \mid \lambda \in \mathbb{C}\} \subset \mathbb{H}$.

(1) With $g_1 : \mathbb{H} \to \mathbb{C}^2$ defined as in Section 2, show that $g_1(\mathbb{C} \cdot q)$ is a one-dimensional $\mathbb{C}$-subspace of $\mathbb{C}^2$.

(2) Define a natural identification $\tilde{g}_1 : \mathbb{H} \to \mathbb{C}^2$ such that $\tilde{g}_1(q \cdot \mathbb{C})$ is a one-dimensional $\mathbb{C}$-subspace of $\mathbb{C}^2$.

Chapter 3

# The orthogonal groups

In this chapter, we define and study what are probably the most important subgroups of the general linear groups. These are denoted $O(n)$, $SO(n)$, $U(n)$, $SU(n)$ and $Sp(n)$. In particular, the group $SO(3)$, which was previously described as the "positions of a globe," now receives a more rigorous definition. We will continue to study these groups throughout the remainder of the book.

## 1. The standard inner product on $\mathbb{K}^n$

The conjugate and norm of an element $q \in \mathbb{K}$ are defined as:

(1) If $q \in \mathbb{R}$, then $\overline{q} = q$ and $|q|$ means the absolute value of $q$.

(2) If $q = a + b\mathbf{i} \in \mathbb{C}$, then $\overline{q} = a - b\mathbf{i}$ and $|q| = \sqrt{a^2 + b^2}$.

(3) If $q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \in \mathbb{H}$, then $\overline{q} = a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}$ and $|q| = \sqrt{a^2 + b^2 + c^2 + d^2}$.

In all cases, it is a quick calculation to verify that for $q, q_1, q_2 \in \mathbb{K}$:

$$\overline{q_1 \cdot q_2} = \overline{q}_2 \cdot \overline{q}_1. \tag{3.1}$$

$$q \cdot \overline{q} = \overline{q} \cdot q = |q|^2. \tag{3.2}$$

These two equalities together imply that:

$$|q_1 \cdot q_2| = |q_1| \cdot |q_2|. \tag{3.3}$$

**Definition 3.1.** *The **standard inner product** on $\mathbb{K}^n$ is the function from $\mathbb{K}^n \times \mathbb{K}^n$ to $\mathbb{K}$ defined by:*

$$\langle (x_1, x_2, ..., x_n), (y_1, y_2, ..., y_n) \rangle_{\mathbb{K}} = x_1 \cdot \overline{y}_1 + x_2 \cdot \overline{y}_2 + \cdots + x_n \cdot \overline{y}_n.$$

It follows from Equation 3.2 that for all $X \in \mathbb{K}^n$, $\langle X, X \rangle_{\mathbb{K}}$ is a real number that is $\geq 0$ and equal to zero only when $X = (0, ..., 0)$. This allows us to define:

**Definition 3.2.** *The **standard norm** on $\mathbb{K}^n$ is the function from $\mathbb{K}^n$ to the nonnegative real numbers defined by:*

$$|X|_{\mathbb{K}} = \sqrt{\langle X, X \rangle_{\mathbb{K}}}.$$

We will omit the $\mathbb{K}$-subscripts whenever there is no ambiguity.

**Proposition 3.3.** *For all $X, Y, Z \in \mathbb{K}^n$ and $\lambda \in \mathbb{K}$,*

(1) $\langle X, Y + Z \rangle = \langle X, Y \rangle + \langle X, Z \rangle$,
(2) $\langle X + Y, Z \rangle = \langle X, Z \rangle + \langle Y, Z \rangle$,
(3) $\langle \lambda X, Y \rangle = \lambda \langle X, Y \rangle$ *and* $\langle X, \lambda Y \rangle = \langle X, Y \rangle \overline{\lambda}$,
(4) $\overline{\langle X, Y \rangle} = \langle Y, X \rangle$.

**Definition 3.4.**

- *Vectors $X, Y \in \mathbb{K}^n$ are called **orthogonal** if $\langle X, Y \rangle = 0$.*
- *A basis $\{X_1, ..., X_n\}$ of $\mathbb{K}^n$ is called **orthonormal** if $\langle X_i, X_j \rangle$ equals $1$ when $i = j$ and equals zero when $i \neq j$ (that is, the vectors have norm $1$ and are mutually orthogonal).*
- *The **standard orthonormal basis** of $\mathbb{K}^n$ is:*

$$e_1 = (1, 0, ..., 0),\ e_2 = (0, 1, 0, ..., 0), ..., e_n = (0, ..., 0, 1).$$

When $\mathbb{K} = \mathbb{R}$, the standard inner product is the familiar "dot product," described geometrically in terms of the angle $\theta$ between $X, Y \in \mathbb{R}^n$:

$$\langle X, Y \rangle_{\mathbb{R}} = |X|_{\mathbb{R}} |Y|_{\mathbb{R}} \cos\theta. \tag{3.4}$$

When $\mathbb{K} = \mathbb{C}$, the standard inner product is also called the **hermitian inner product**. Since the hermitian inner product of two vectors $X, Y \in \mathbb{C}^n$ is a complex number, we should separately interpret the geometric meanings of its real and imaginary parts. The cleanest such interpretation is in terms of the identification

$$f = f_n : \mathbb{C}^n \to \mathbb{R}^{2n}$$

from the previous chapter. It is easy to verify that for all $X, Y \in \mathbb{C}^n$,

$$\langle X, Y\rangle_{\mathbb{C}} = \langle f(X), f(Y)\rangle_{\mathbb{R}} + \mathbf{i}\langle f(X), f(\mathbf{i}Y)\rangle_{\mathbb{R}}, \tag{3.5}$$

$$|X|_{\mathbb{C}} = |f(X)|_{\mathbb{R}}. \tag{3.6}$$

Thus, if $X, Y \in \mathbb{C}^n$ are orthogonal, then two things are true:

$$\langle f(X), f(Y)\rangle_{\mathbb{R}} = 0 \quad \text{and} \quad \langle f(X), f(\mathbf{i}Y)\rangle_{\mathbb{R}} = 0.$$

This observation leads to:

**Proposition 3.5.** $\{X_1, ..., X_n\} \in \mathbb{C}^n$ *is an orthonormal basis if and only if* $\{f(X_1), f(\mathbf{i}X_1), ..., f(X_n), f(\mathbf{i}X_n)\}$ *is an orthonormal basis of* $\mathbb{R}^{2n}$.

When $\mathbb{K} = \mathbb{H}$, the standard inner product is also called the **symplectic inner product**. For $X, Y \in \mathbb{H}^n$, the $1$, $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ components of $\langle X, Y\rangle_{\mathbb{H}}$ are best interpreted geometrically in terms of the identification $h = f_{2n} \circ g_n : \mathbb{H}^n \to \mathbb{R}^{4n}$, as follows:

$$\begin{aligned} \langle X, Y\rangle_{\mathbb{H}} &= \langle h(X), h(Y)\rangle_{\mathbb{R}} + \mathbf{i}\langle h(X), h(\mathbf{i}Y)\rangle_{\mathbb{R}} \\ &\quad + \mathbf{j}\langle h(X), h(\mathbf{j}Y)\rangle_{\mathbb{R}} + \mathbf{k}\langle h(X), h(\mathbf{k}Y)\rangle_{\mathbb{R}}. \\ |X|_{\mathbb{H}} &= |h(X)|_{\mathbb{R}}. \end{aligned}$$

**Proposition 3.6.** $\{X_1, ..., X_n\} \in \mathbb{H}^n$ *is an orthonormal basis if and only if the following is an orthonormal basis of* $\mathbb{R}^{4n}$:

$$\{h(X_1), h(\mathbf{i}X_1), h(\mathbf{j}X_1), h(\mathbf{k}X_1), ..., h(X_n), h(\mathbf{i}X_n), h(\mathbf{j}X_n), h(\mathbf{k}X_n)\}.$$

The following inequality follows from Equation 3.4 when $\mathbb{K} = \mathbb{R}$:

**Proposition 3.7** (Schwarz inequality)**.** *For all* $X, Y \in \mathbb{K}^n$,

$$|\langle X, Y\rangle| \leq |X| \cdot |Y|.$$

**Proof.** Let $X, Y \in \mathbb{K}^n$. Let $\alpha = \langle X, Y\rangle$. Assume that $X \neq 0$ (otherwise the proposition is trivial). For all $\lambda \in \mathbb{K}$, we have:

$$\begin{aligned} 0 &\leq |\lambda X + Y|^2 = \langle \lambda X + Y, \lambda X + Y\rangle \\ &= \lambda\langle X, X\rangle\overline{\lambda} + \lambda\langle X, Y\rangle + \langle Y, X\rangle\overline{\lambda} + \langle Y, Y\rangle \\ &= |\lambda|^2|X|^2 + \lambda\langle X, Y\rangle + \overline{\lambda\langle X, Y\rangle} + |Y|^2 \\ &= |\lambda|^2|X|^2 + 2\mathrm{Re}(\lambda\alpha) + |Y|^2. \end{aligned}$$

Choosing $\lambda = -\overline{\alpha}/|X|^2$ gives:

$$0 \leq |\alpha|^2/|X|^2 - 2|\alpha|^2/|X|^2 + |Y|^2,$$

which proves that $|\alpha| \leq |X| \cdot |Y|$ as desired. □

## 2. Several characterizations of the orthogonal groups

**Definition 3.8.** *The* ***orthogonal group over*** $\mathbb{K}$,

$$\mathcal{O}_n(\mathbb{K}) = \{A \in GL_n(\mathbb{K}) \mid \langle X \cdot A, Y \cdot A\rangle = \langle X, Y\rangle \text{ for all } X, Y \in \mathbb{K}^n\},$$

... *is denoted $O(n)$ and called the* ***orthogonal group*** *if $\mathbb{K} = \mathbb{R}$.*

... *is denoted $U(n)$ and called the* ***unitary group*** *if $\mathbb{K} = \mathbb{C}$.*

... *is denoted $Sp(n)$ and called the* ***symplectic group*** *if $\mathbb{K} = \mathbb{H}$.*

It is straightforward to see that $\mathcal{O}_n(\mathbb{K})$ is a subgroup of $GL_n(\mathbb{K})$. Its elements are called **orthogonal**, **unitary** or **symplectic** matrices. To describe their form, it is useful to denote the **conjugate-transpose** of $A \in M_n(\mathbb{K})$ as $A^* = (\overline{A})^T$, where $\overline{A}$ means the matrix obtained by conjugating all of the entries of $A$.

**Proposition 3.9.** *For $A \in GL_n(\mathbb{K})$ the following are equivalent.*

(1) *$A \in \mathcal{O}_n(\mathbb{K})$.*

(2) *$R_A$ preserves orthonormal bases; i.e., if $\{X_1, ..., X_n\}$ is an orthonormal basis of $\mathbb{K}^n$, then so is $\{R_A(X_1), ..., R_A(X_n)\}$.*

(3) *The rows of $A$ form an orthonormal basis of $\mathbb{K}^n$.*

(4) *$A \cdot A^* = I$.*

**Proof.** (1) $\implies$ (2) is obvious. (2) $\implies$ (3) because the rows of $A$ equal $\{R_A(e_1), ..., R_A(e_n)\}$. To see that (3) $\iff$ (4), notice that:

$$\begin{aligned}(A \cdot A^*)_{ij} &= (\text{row } i \text{ of } A) \cdot (\text{column } j \text{ of } A^*) \\ &= (\text{row } i \text{ of } A) \cdot (\text{row } j \text{ of } \overline{A})^T \\ &= \langle (\text{row } i \text{ of } A), (\text{row } j \text{ of } A) \rangle.\end{aligned}$$

Finally, we prove that (3) $\implies$ (1). If the rows of $A$ are orthonormal, then for all $X = (x_1, ..., x_n), Y = (y_1, ..., y_n) \in \mathbb{K}^n$,

$$\begin{aligned}&\langle R_A(X), R_A(Y) \rangle \\ &\quad = \left\langle \sum_{l=1}^{n} x_l (\text{row } l \text{ of } A), \sum_{s=1}^{n} y_s (\text{row } s \text{ of } A) \right\rangle \\ &\quad = \sum_{l,s=1}^{n} x_l \langle (\text{row } l \text{ of } A), (\text{row } s \text{ of } A) \rangle \overline{y}_s \\ &\quad = x_1\overline{y}_1 + \cdots + x_n\overline{y}_n = \langle X, Y \rangle.\end{aligned}$$

□

Geometrically, $O(n)$ is the group of matrices $A$ for which the linear transformation $R_A : \mathbb{R}^n \to \mathbb{R}^n$ preserves dot products of vectors, and hence also norms of vectors. Such transformations should be visualized as "rigid motions" of $\mathbb{R}^n$ (we will be more precise about this in Section 5). The geometric meanings of $U(n)$ and $Sp(n)$ are best described in terms $O(n)$ by considering the homomorphisms from the previous chapter.

**Proposition 3.10.**

(1) $\rho_n(U(n)) = O(2n) \cap \rho_n(GL_n(\mathbb{C}))$.
(2) $\Psi_n(Sp(n)) = U(2n) \cap \Psi_n(GL_n(\mathbb{H}))$.
(3) $(\rho_{2n} \circ \Psi_n)(Sp(n)) = O(4n) \cap (\rho_{2n} \circ \Psi_n)(GL_n(\mathbb{H}))$.

Since $U(n)$ is isomorphic to its image, $\rho_n(U(n))$, part (1) says that $U(n)$ is isomorphic to the group of complex-linear real orthogonal matrices. In other words, $U(n)$ is isomorphic to the group of rigid motions of $\mathbb{R}^{2n}$ that preserve the standard complex structure. Similarly, part (3) says that $Sp(n)$ is isomorphic to the group of quaternionic-linear real orthogonal matrices.

**Proof.** We prove only (1), since (2) is similar and (3) follows from (1) and (2). The most straightforward idea is to use Equation 3.5. But a quicker approach is to first notice that for all $A \in M_n(\mathbb{C})$,

$$\rho_n(A^*) = \rho_n(A)^*.$$

If $A \in GL_n(\mathbb{C})$, then $\rho_n(A) \cdot \rho_n(A)^* = \rho_n(A) \cdot \rho_n(A^*) = \rho_n(A \cdot A^*)$, which shows that $A \in U(n)$ if and only if $\rho_n(A) \in O(2n)$. □

We said that $\mathcal{O}_n(\mathbb{K})$ is the group of matrices $A$ for which $R_A$ preserves inner products of vectors, and hence also norms of vectors. The next result says that if $R_A$ preserves norms, then it automatically preserves inner products.

**Proposition 3.11.**

$$\mathcal{O}_n(\mathbb{K}) = \{A \in GL_n(\mathbb{K}) \mid |R_A(X)| = |X| \textit{ for all } X \in \mathbb{K}^n\}.$$

**Proof.** To prove the case $\mathbb{K} = \mathbb{R}$, we show that the inner product is completely determined by the norm. Solving the equation

$$|X - Y|_{\mathbb{R}}^2 = \langle X - Y, X - Y\rangle_{\mathbb{R}} = \langle X, X\rangle_{\mathbb{R}} + \langle Y, Y\rangle_{\mathbb{R}} - 2\langle X, Y\rangle_{\mathbb{R}}$$

for $\langle X, Y\rangle_{\mathbb{R}}$ gives:

$$\langle X, Y\rangle_{\mathbb{R}} = \frac{1}{2}\left(|X|_{\mathbb{R}}^2 + |Y|_{\mathbb{R}}^2 - |X - Y|_{\mathbb{R}}^2\right). \tag{3.7}$$

From this, it is straightforward to show that if $R_A$ preserves norms, then it also preserves inner products.

The above argument doesn't work for $\mathbb{K} \in \{\mathbb{C}, \mathbb{H}\}$ (why not?). Instead, we prove the case $\mathbb{K} = \mathbb{C}$ as a consequence of the real case. Suppose $A \in GL_n(\mathbb{C})$ is such that $R_A : \mathbb{C}^n \to \mathbb{C}^n$ is norm-preserving. Then $R_{\rho_n(A)} : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ also preserves norms, since for all $X \in \mathbb{C}^n$,

$$|R_{\rho_n(A)}(f_n(X))|_{\mathbb{R}} = |f_n(R_A(X))|_{\mathbb{R}} = |R_A(X)|_{\mathbb{C}} = |X|_{\mathbb{C}} = |f_n(X)|_{\mathbb{R}}.$$

Therefore $\rho_n(A) \in O(2n)$, which using Proposition 3.10 implies that $A \in U(n)$.

The $\mathbb{K} = \mathbb{H}$ case is proven from the real case in a similar fashion. □

## 3. The special orthogonal groups

In this section, we define important subgroups of the orthogonal groups, beginning with the observation that:

**Proposition 3.12.** *If* $A \in \mathcal{O}_n(\mathbb{K})$*, then* $|\det(A)| = 1$.

**Proof.** Since $A \cdot A^* = I$,

$$1 = \det(A \cdot A^*) = \det(A) \cdot \det(A^*) = \det(A) \cdot \overline{\det(A)} = |\det(A)|^2.$$

We used the fact that $\det(A^*) = \overline{\det(A)}$, which should be verified first for $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. The quaternionic case follows from the complex case because for quaternionic matrices, $\det(A)$ means $\det(\Psi_n(A))$, and $\Psi_n(A^*) = \Psi_n(A)^*$. ☐

The interpretation of Proposition 3.12 depends on $\mathbb{K}$:

- If $A \in O(n)$, then $\det(A) = \pm 1$.
- If $A \in U(n)$, then $\det(A) = e^{i\theta}$ for some $\theta \in [0, 2\pi)$.
- If $A \in Sp(n)$, then Proposition 2.10 implies $\det(A) = \pm 1$. We will see later that $\det(A) = 1$.

The subgroup

$$SO(n) = \{A \in O(n) \mid \det(A) = 1\}$$

is called the **special orthogonal group**. The subgroup

$$SU(n) = \{A \in U(n) \mid \det(A) = 1\}$$

is called the **special unitary group**. Both are clearly subgroups of the general linear group and in fact of the **special linear group**:

$$SL_n(\mathbb{K}) = \{A \in GL_n(\mathbb{K}) \mid \det(A) = 1\}.$$

Notice that $SO(n)$ comprises the orthogonal matrices whose determinants are one of two possibilities, while $SU(n)$ comprises the unitary matrices whose determinants are one of a circle's worth of possibilities. We will see later that the relationship of $SO(n)$ to $O(n)$ is very different from $SU(n)$ to $U(n)$.

## 4. Low dimensional orthogonal groups

In this section, we explicitly describe $\mathcal{O}_n(\mathbb{K})$ for small values of $n$. First, $O(1) = \{(1), (-1)\}$ and $SO(1) = \{(1)\}$ are isomorphic to the unique groups with 2 and 1 elements respectively.

Next, if $A \in O(2)$, then its two rows form an orthonormal basis of $\mathbb{R}^2$. Its first row is an arbitrary unit-length vector of $\mathbb{R}^2$, which can be written as $(\cos\theta, \sin\theta)$ for some $\theta$. The second row is unit-length and orthogonal to the first, which leaves two choices: $(-\sin\theta, \cos\theta)$ or $(\sin\theta, -\cos\theta)$. For the first choice, $\det(A) = 1$, and for the second, $\det(A) = -1$. So we learn:

$$SO(2) = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \,\middle|\, \theta \in [0, 2\pi) \right\}, \tag{3.8}$$

$$O(2) = SO(2) \cup \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \,\middle|\, \theta \in [0, 2\pi) \right\}.$$

$SO(2)$ is identified with the set of points on a circle; its group operation is addition of angles. $O(2)$ is a disjoint union of two circles. It is interesting that the disjoint union of two circles has a group operation.

Next, $SU(1) = \{(1)\}$ and $U(1) = \{(e^{i\theta}) \mid \theta \in [0, 2\pi)\}$, which is isomorphic to the circle-group $SO(2)$.

Next, $Sp(1) = \{(a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) \mid a^2 + b^2 + c^2 + d^2 = 1\}$ is the **group of unit-length quaternions**, which is naturally identified with the three-dimensional sphere $S^3 \subset \mathbb{R}^4 \cong \mathbb{H}$. In fact, it follows from Equation 3.3 that the product of two unit-length quaternions is a unit-length quaternion. So we might have mentioned several pages ago the beautiful fact that *quaternionic multiplication provides a group operation on the three-dimensional sphere!* It turns out that $S^0$, $S^1$ and $S^3$ are the only spheres which are also groups.

We conclude this section by showing that $SU(2)$ is isomorphic to $Sp(1)$, and thus in some sense also has the shape of a three-dimensional sphere.

**Proposition 3.13.** *$SU(2)$ is isomorphic to $Sp(1)$.*

**Proof.** First notice that

$$\Psi_1(Sp(1)) = \left\{ \begin{pmatrix} z & w \\ -\overline{w} & \overline{z} \end{pmatrix} \mid z, w \in \mathbb{C} \text{ such that } |z|^2 + |w|^2 = 1 \right\}$$

is a subgroup of $U(2)$ by Proposition 3.10, namely, the quaternionic-linear 2-by-2 unitary matrices. Calculating the determinant of such matrices shows that $\Psi_1(Sp(1)) \subset SU(2)$. We wish to prove that $\Psi_1(Sp(1)) = SU(2)$, so that $\Psi_1$ determines an isomorphism between $Sp(1)$ and $SU(2)$.

Let $A = \begin{pmatrix} z_1 & w_1 \\ w_2 & z_2 \end{pmatrix} \in SU(2)$. An easily verified formula for the inverse of a 2-by-2 matrix is: $A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} z_2 & -w_1 \\ -w_2 & z_1 \end{pmatrix}$. In our case, $\det(A) = 1$, so $\begin{pmatrix} z_2 & -w_1 \\ -w_2 & z_1 \end{pmatrix} = A^{-1} = A^* = \begin{pmatrix} \overline{z}_1 & \overline{w}_2 \\ \overline{w}_1 & \overline{z}_2 \end{pmatrix}$, which tells us that $z_2 = \overline{z}_1$ and $w_2 = -\overline{w}_1$. It now follows that $SU(2) = \Psi_1(Sp(1))$. □

## 5. Orthogonal matrices and isometries

In this section, we describe $O(n)$ geometrically as the group of isometries of $\mathbb{R}^n$ that fix the origin and we discuss the difference between $SO(3)$ and $O(3)$.

The **distance** between points $X = (x_1, ..., x_n)$ and $Y = (y_1, ..., y_n)$ in $\mathbb{R}^n$ is measured as:

$$\text{dist}(X, Y) = |X - Y| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

A function $f : \mathbb{R}^n \to \mathbb{R}^n$ is called an **isometry** if for all $X, Y \in \mathbb{R}^n$, $\text{dist}(f(X), f(Y)) = \text{dist}(X, Y)$ .

**Proposition 3.14.**

(1) *If $A \in O(n)$, then $R_A : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry.*

(2) *If $f : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry with $f(0) = 0$, then $f = R_A$ for some $A \in O(n)$. In particular, $f$ is linear.*

**Proof.** For part (1), if $A \in O(n)$, then for all $X, Y \in \mathbb{R}^n$,

$$\begin{aligned} \text{dist}(R_A(X), R_A(Y)) &= |R_A(X) - R_A(Y)| = |R_A(X - Y)| \\ &= |X - Y| = \text{dist}(X, Y), \end{aligned}$$

which proves that $R_A$ is an isometry.

For part (2), suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry for which $f(0) = 0$. Equation 3.7 can be re-expressed as a description of the inner product completely in terms of distances:

$$\langle X, Y \rangle = \frac{1}{2}\left(\text{dist}(X, 0)^2 + \text{dist}(Y, 0)^2 - \text{dist}(X, Y)^2\right).$$

Since $f$ preserves distances and fixes the origin, it is straightforward to show using this that it must also preserve the inner product:

$$\langle f(X), f(Y) \rangle = \langle X, Y \rangle \text{ for all } X, Y \in \mathbb{R}^n.$$

Let $A$ be the matrix whose $i^{\text{th}}$ row is $f(e_i)$, so $f(e_i) = R_A(e_i)$ for all $i = 1, ..., n$. Notice that $A \in O(n)$, since its rows are orthonormal. We will prove that $f = R_A$ (and thus that $f$ is linear) by showing that $g = (R_A)^{-1} \circ f$ is the identity function. Notice that $g$ is an isometry with $g(0) = 0$ (so $g$ preserves inner products, as above) and $g(e_i) = e_i$ for all $i = 1, ..., n$. Let $X \in \mathbb{R}^n$. Write $X = \sum a_i e_i$ and $g(X) = \sum b_i e_i$. Then,

$$b_i = \langle g(X), e_i \rangle = \langle g(X), g(e_i) \rangle = \langle X, e_i \rangle = a_i,$$

which proves $g(X) = X$, so $g$ is the identity function. $\square$

$O(n)$ is the group of isometries of $\mathbb{R}^n$ that fix the origin and that therefore map the sphere $S^{n-1} \subset \mathbb{R}^n$ to itself. For example, elements of $O(3)$ represent functions from the "globe" $S^2 \subset \mathbb{R}^3$ to itself. We will see next that elements of $SO(3)$ represent real physical motions of the globe, which justifies our characterization of $SO(3)$ as the group of positions of a globe (Section 1 of Chapter 1).

To understand the difference between $O(3)$ and $SO(3)$, we must discuss the **orientation** of $\mathbb{R}^3$. An ordered orthonormal basis of $\mathbb{R}^3$, like $\{X_1, X_2, X_3\}$, is called **right-handed** if $X_1 \times X_2 = X_3$, where "$\times$" denotes the vector cross product in $\mathbb{R}^3$. Visually, this means that if the fingers of your right hand are curled from $X_1$ towards $X_2$, then your thumb will point in the direction of $X_3$.

**Proposition 3.15.** *Let $A \in O(3)$. Then $A \in SO(3)$ if and only if the rows of $A$, $\{R_A(e_1), R_A(e_2), R_A(e_3)\}$, form a right-handed ordered orthonormal basis.*

**Proof.** Let $R_A(e_1) = (a, b, c)$ and $R_A(e_2) = (d, e, f)$ denote the first two rows of $A$. The third row is unit-length and orthogonal to both, which leaves two choices:

$$R_A(e_3) = \pm(R_A(e_1) \times R_A(e_2)) = \pm(bf - ce, cd - af, ae - bd).$$

A quick calculation shows that the "+" choice gives $\det(A) > 0$, while the "−" choice gives $\det(A) < 0$. □

Elements of $SO(3)$ correspond to "physically performable motions" of a globe. This statement is imprecise, but in Chapter 9 we give it teeth by proving that every element of $SO(3)$ is a rotation through some angle about some single axis. An element of $O(3)$ with negative determinant turns the globe inside-out. For example, $R_{\mathrm{diag}(-1,-1,-1)}$ maps each point of the globe to its antipode (its negative). This is not a physically performable motion.

## 6. The isometry group of Euclidean space

It is a straightforward exercise to show that

$$\mathrm{Isom}(\mathbb{R}^n) = \{f : \mathbb{R}^n \to \mathbb{R}^n \mid f \text{ is an isometry}\}$$

is a group under composition of functions. The subgroup of isometries that fix the origin is isomorphic to $O(n)$. An isometry, $f$, that does not fix the origin is not linear, so cannot equal $R_A$ for any matrix $A$. In this case, let $V = f(0)$, so the function $X \mapsto f(X) - V$ is an isometry that fixes the origin and therefore equals $R_A$ for some $A \in O(n)$. Therefore, an arbitrary isometry of $\mathbb{R}^n$ has the form

$$f(X) = R_A(X) + V$$

for some $A \in O(n)$ and $V \in \mathbb{R}^n$.

There is a clever trick for representing any isometry of $\mathbb{R}^n$ as a matrix, even ones that do not fix the origin. Graphics programmers use this trick to rotate *and translate* objects on the computer screen via matrices. We first describe the $n = 3$ case.

Let $A \in O(3)$ and $V = (v_1, v_2, v_3) \in \mathbb{R}^3$. We will represent the isometry $f(X) = R_A(X) + V$ by the matrix:

$$F = \begin{pmatrix} A & 0 \\ V & 1 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} & 0 \\ A_{21} & A_{22} & A_{23} & 0 \\ A_{31} & A_{32} & A_{33} & 0 \\ v_1 & v_2 & v_3 & 1 \end{pmatrix} \in GL_4(\mathbb{R}).$$

Let $X = (x_1, x_2, x_3) \in \mathbb{R}^3$. Denote $(X, 1) = (x_1, x_2, x_3, 1) \in \mathbb{R}^4$. Notice that

$$(X, 1) \cdot F = (f(X), 1) \in \mathbb{R}^4.$$

In this way, $F$ represents $f$.

The composition of two isometries, like the ones represented by $F_1 = \begin{pmatrix} A_1 & 0 \\ V_1 & 1 \end{pmatrix}$ and $F_2 = \begin{pmatrix} A_2 & 0 \\ V_2 & 1 \end{pmatrix}$, is the isometry represented by the product:

$$\begin{pmatrix} A_1 & 0 \\ V_1 & 1 \end{pmatrix} \cdot \begin{pmatrix} A_2 & 0 \\ V_2 & 1 \end{pmatrix} = \begin{pmatrix} A_1 \cdot A_2 & 0 \\ R_{A_2}(V_1) + V_2 & 1 \end{pmatrix}.$$

Matrix multiplication is quite useful here. It allowed us to see immediately that the isometry $X \mapsto R_{A_1}(X) + V_1$ followed by the isometry $X \mapsto R_{A_2}(X) + V_2$ is the isometry $X \mapsto R_{(A_1 \cdot A_2)}(X) + R_{A_2}(V_1) + V_2$.

The above ideas also work for values of $n$ other than 3. We conclude that $\text{Isom}(\mathbb{R}^n)$ is isomorphic to the following subgroup of $GL_{n+1}(\mathbb{R})$:

$$\text{Isom}(\mathbb{R}^n) \cong \left\{ \begin{pmatrix} A & 0 \\ V & 1 \end{pmatrix} \,\middle|\, A \in O(n) \text{ and } V \in \mathbb{R}^n \right\}.$$

Notice that the following subgroup of $\text{Isom}(\mathbb{R}^n)$ is isomorphic to $(\mathbb{R}^n, +)$, which denotes $\mathbb{R}^n$ under the group-operation of vector-addition:

$$\text{Trans}(\mathbb{R}^n) = \left\{ \begin{pmatrix} I & 0 \\ V & 1 \end{pmatrix} \,\middle|\, V \in \mathbb{R}^n \right\}.$$

This is the group of isometries of $\mathbb{R}^n$ that only translate and do not rotate. It is interesting that $(\mathbb{R}^n, +)$ is isomorphic to a matrix group!

## 7. Symmetry groups

The **symmetry group** of a subset $X \subset \mathbb{R}^n$ is the group of all isometries of $\mathbb{R}^n$ that carry $X$ onto itself:

**Definition 3.16.** *$Symm(X) = \{f \in Isom(\mathbb{R}^n) \mid f(X) = X\}$.*

The statement "$f(X) = X$" means that each point of $X$ is sent by $f$ to a (possibly different) point of $X$.

For example, the symmetry group of the sphere $S^{n-1} \subset \mathbb{R}^n$ equals the group of isometries of $\mathbb{R}^n$ with no translational component, which is isomorphic to the orthogonal group:

$$\mathrm{Symm}(S^{n-1}) = \left\{ \begin{pmatrix} A & 0 \\ V & 1 \end{pmatrix} \middle| A \in O(n),\, V = (0, ..., 0) \right\} \cong O(n).$$

In an abstract algebra course, you probably encountered some important finite symmetry groups. For example, the symmetry group of a regular $m$-gon (triangle, square, pentagon, hexagon, etc.) centered at the origin in $\mathbb{R}^2$ is called the **dihedral group** of order $2m$, denoted $D_m$. The elements of $D_m$ with determinant $+1$ are called rotations; they form a subgroup of index 2 that is isomorphic to the cyclic group $\mathbb{Z}_m$, of order $m$. The elements of $D_m$ with determinant $-1$ are called flips.

The fact that half of the elements of $D_m$ are rotations illustrates a general principal:

**Definition 3.17.** *$Symm(X) = Symm^+(X) \cup Symm^-(X)$, where the sets*

$$Symm^{\pm}(X) = \left\{ \begin{pmatrix} A & 0 \\ V & 1 \end{pmatrix} \in Symm(X) \middle| \det(A) = \pm 1 \right\}$$

*are called the **proper** and **improper** symmetry groups of $X$.*

**Proposition 3.18.** *For any $X \subset \mathbb{R}^n$, $Symm^+(X) \subset Symm(X)$ is a subgroup with index 1 or 2.*

The proof is left to the reader in Exercise 3.4. An example of a set $Y \subset \mathbb{R}^2$ whose proper symmetry group has index 1 (meaning all symmetries are proper) is illustrated in Figure 1 (right).

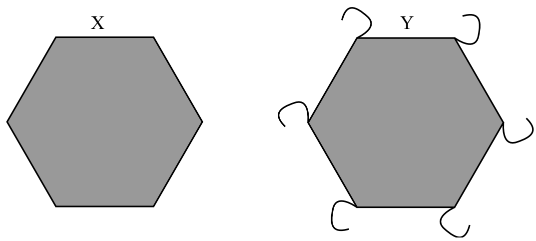



**Figure 1.** $\mathrm{Symm}(X) = D_6$, while $\mathrm{Symm}(Y) = \mathbb{Z}_6$.

Symmetry groups of subsets of $\mathbb{R}^2$ are useful for studying objects that are essentially 2-dimensional, like snowflakes and certain crystal structures. Many subsets of $\mathbb{R}^2$, like the wallpaper tilings of $\mathbb{R}^2$ illustrated in some M.C. Escher prints, have infinite symmetry groups. Chapter 28 of [**5**] describes the classification of such infinite "wallpaper groups". Perhaps surprisingly, the only *finite* symmetry groups in dimension 2 are $D_m$ and $\mathbb{Z}_m$. The following theorem is attributed to Leonardo da Vinci (1452-1519):

**Proposition 3.19.** *For $X \subset \mathbb{R}^2$, if Symm$(X)$ is finite, then it is isomorphic to $D_m$ or $\mathbb{Z}_m$ for some $m$.*

The proof involves two steps. First, when $\mathrm{Symm}(X)$ is finite, its elements must share a common fixed point, so it is isomorphic to a subgroup of $O(2)$. Second, $D_m$ and $\mathbb{Z}_m$ are the only finite subgroups of $O(2)$.

Symmetry groups of subsets of $\mathbb{R}^3$ are even more interesting. In chemistry, the physical properties of a substance are intimately related to the symmetry groups of its molecules. In dimension 3, there are still very few possible finite symmetry groups:

**Theorem 3.20.** *For $X \subset \mathbb{R}^3$, if Symm$^+(X)$ is finite, then it is isomorphic to $D_m$, $\mathbb{Z}_m$, $A_4$, $S_4$ or $A_5$.*

Here, $S_m$ denotes the group of permutations of a set with $m$ elements, and $A_m \subset S_m$ denotes the subgroup of even permutations (called the **alternating group**). Like the $n = 2$ case, the proof involves verifying that all symmetries have a common fixed point and that the only finite subgroups of $SO(3)$ are $D_m$, $\mathbb{Z}_m$, $A_4$, $S_4$ and $A_5$.

The *regular solids* provide examples of sets whose proper symmetry groups equal $A_4$, $S_4$ and $A_5$. A **regular solid** (also called a "platonic solid" or a "regular polyhedra") is a polyhedra whose faces are mutually congruent regular polygons, at each of whose vertices the same number of edges meet. A famous classification theorem, attributed to Plato around 400 B.C., says that there are only five regular solids, pictured in Figure 2. The regular solids were once con-

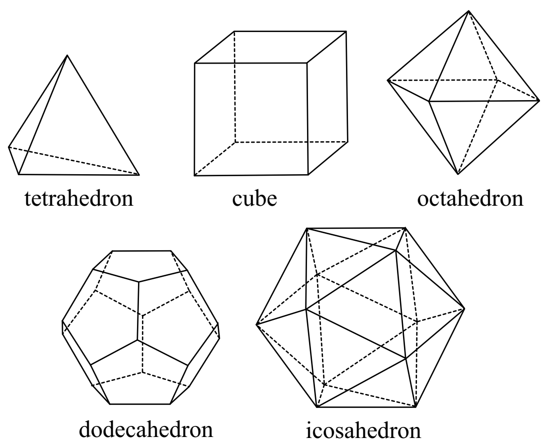



**Figure 2.** The five regular solids.

sidered to be sacred shapes, thought to represent fire, earth, air, the universe, and water.

The regular solids exemplify the last three possibilities enumerated in Theorem 3.20, which enhances one's sense that they are of universal importance. It turns out that $A_4$ is the proper symmetry group of a tetrahedron, $S_4$ is the proper symmetry group of a cube or an octahedron, and $A_5$ is the proper symmetry group of a dodecahedron or an icosahedron. See [**6**] for a complete calculation of these proper symmetry groups and a proof of Theorem 3.20. Since a cube has 6 faces, 12 edges, and 8 vertices, it may be surprising that its proper symmetry group is $S_4$. What does a cube have 4 of which get permuted by its proper symmetries? It has 4 diagonals (lines connecting antipodal pairs of vertices). This observation is the starting point of the calculation of its proper symmetry group.

## 8. Exercises

**Ex. 3.1.** Prove part (4) of Proposition 3.3.

**Ex. 3.2.** Prove Equations 3.5 and 3.6.

**Ex. 3.3.** Prove Proposition 3.5.

**Ex. 3.4.** Prove Proposition 3.18.

**Ex. 3.5.** Let $A \in GL_n(\mathbb{K})$. Prove that $A \in \mathcal{O}_n(\mathbb{K})$ if and only if the columns of $A$ are an orthonormal basis of $\mathbb{K}^n$.

**Ex. 3.6.**

(1) Show that for every $A \in O(2) - SO(2)$, $R_A : \mathbb{R}^2 \to \mathbb{R}^2$ is a flip about some line through the origin. How is this line determined by the angle of $A$ (as in Equation 3.8)?

(2) Let $B = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \in SO(2)$. Assume that $\theta$ is not an integer multiple of $\pi$. Prove that $B$ does not commute with any $A \in O(2) - SO(2)$. *Hint: Show that $R_{AB}$ and $R_{BA}$ act differently on the line in $\mathbb{R}^2$ about which $A$ is a flip.*

**Ex. 3.7.** Describe the product of two arbitrary elements of $O(2)$ in terms of their angles (as in Equation 3.8).

**Ex. 3.8.** Let $A \in O(n)$ have determinant $-1$. Prove that:

$$O(n) = SO(n) \cup \{A \cdot B \mid B \in SO(n)\}.$$

**Ex. 3.9.** Define a map $f : O(n) \to SO(n) \times \{+1, -1\}$ as follows:

$$f(A) = (\det(A) \cdot A, \det A).$$

(1) If $n$ is odd, prove that $f$ is an isomorphism.

(2) Assume that $n$ is odd and that $X \subset \mathbb{R}^n$ is symmetric about the origin, which means that $-p \in X$ if and only if $p \in X$. Also assume that $\mathrm{Symm}(X) \subset O(n)$; in other words, $X$ has no translational symmetries. Prove that $\mathrm{Symm}(X)$ is isomorphic to $\mathrm{Symm}^+(X) \times \{+1, -1\}$.
*Comment: Four of the five regular solids are symmetric about the origin. The tetrahedron is not; its proper symmetry group is $A_4$ and its full symmetry group is $S_4$.*

(3) Prove that $O(2)$ is not isomorphic to $SO(2) \times \{+1, -1\}$. *Hint: How many elements of order two are there?*

**Ex. 3.10.** Prove that $\text{Trans}(\mathbb{R}^n)$ is a normal subgroup of $\text{Isom}(\mathbb{R}^n)$.

**Ex. 3.11.** Prove that the **Affine group**,

$$\text{Aff}_n(\mathbb{K}) = \left\{ \begin{pmatrix} A & 0 \\ V & 1 \end{pmatrix} \middle| A \in GL_n(\mathbb{K}) \text{ and } V \in \mathbb{K}^n \right\},$$

is a subgroup of $GL_{n+1}(\mathbb{K})$. Any $F \in \text{Aff}_n(\mathbb{K})$ can be identified with the function $f(X) = R_A(X) + V$ from $\mathbb{K}^n$ to $\mathbb{K}^n$ as in Section 6. Prove that $f$ sends translated lines in $\mathbb{K}^n$ to translated lines in $\mathbb{K}^n$. A *translated line* in $\mathbb{K}^n$ means a set of the form $\{v_0 + v | v \in W\}$, where $v_0 \in \mathbb{K}^n$, and $W \subset \mathbb{K}^n$ is a 1-dimensional $\mathbb{K}$-subspace.

**Ex. 3.12.** Is $\text{Aff}_1(\mathbb{R})$ abelian? Explain algebraically and visually.

**Ex. 3.13.** Let $A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$.

(1) Calculate $R_A(x, y, z, w)$.

(2) Describe a subgroup, $H$, of $O(4)$ that is isomorphic to $S_4$ ($S_4 =$ the group of permutations of a 4 element set).

(3) Describe a subgroup, $H$, of $O(n)$ that is isomorphic to $S_n$. What is $H \cap SO(n)$?

(4) Prove that every finite group is isomorphic to a subgroup of $O(n)$ for some integer $n$. *Hint: Use Cayley's Theorem, found in any abstract algebra textbook.*

**Ex. 3.14.** Let $\mathfrak{g}$ be a $\mathbb{K}$-subspace of $\mathbb{K}^n$ with dimension $d$. Let $\mathcal{B} = \{X_1, ..., X_d\}$ be an orthonormal basis of $\mathfrak{g}$. Let $f : \mathfrak{g} \to \mathfrak{g}$ be $\mathbb{K}$-linear. Let $A \in M_d(\mathbb{K})$ represent $f$ in the basis $\mathcal{B}$. Prove that the following are equivalent:

(1) $A \in \mathcal{O}_d(\mathbb{K})$.

(2) $\langle f(X), f(Y) \rangle = \langle X, Y \rangle$ for all $X, Y \in \mathfrak{g}$.

Show by example that this is false when $\mathcal{B}$ is not orthonormal.

**Ex. 3.15.** Prove that the tetrahedron's symmetry group is $S_4$ and its proper symmetry group is $A_4$. *Hint: The symmetries permute the four vertices.*

**Ex. 3.16.** Think of $Sp(1)$ as the group of unit-length quaternions; that is, $Sp(1) = \{q \in \mathbb{H} \mid |q| = 1\}$.

(1) For every $q \in Sp(1)$, show that the conjugation map $C_q : \mathbb{H} \to \mathbb{H}$, defined as $C_q(V) = q \cdot V \cdot \overline{q}$, is an orthogonal linear transformation. Thus, with respect to the natural basis $\{1, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$ of $\mathbb{H}$, $C_q$ can be regarded as an element of $O(4)$.

(2) For every $q \in Sp(1)$, verify that $C_q(1) = 1$ and therefore that $C_q$ sends $\mathrm{Im}(\mathbb{H}) = \mathrm{span}\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ to itself. Conclude that the restriction $C_q\big|_{\mathrm{Im}(\mathbb{H})}$ can be regarded as an element of $O(3)$.

(3) Define $\varphi : Sp(1) \to O(3)$ as:
$$\varphi(q) = C_q\big|_{\mathrm{Im}(\mathbb{H})}.$$
Verify that $\varphi$ is a group homomorphism.

(4) Verify that the kernel of $\varphi$ is $\{1, -1\}$ and therefore that $\varphi$ is two-to-one.

*Comment: We will show later that the image of $\varphi$ is $SO(3)$.*

**Ex. 3.17.** Think of $Sp(1) \times Sp(1)$ as the group of pairs of unit-length quaternions.

(1) For every $q = (q_1, q_2) \in Sp(1) \times Sp(1)$, show that the map $F(q) : \mathbb{H} \to \mathbb{H}$ defined as $F(q)(V) = q_1 \cdot V \cdot \overline{q_2}$, is an orthogonal linear transformation. Thus, with respect to the natural basis $\{1, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$ of $\mathbb{H}$, $F(q)$ can be regarded as an element of $O(4)$.

(2) Show that this function $F : Sp(1) \times Sp(1) \to O(4)$ is a group homomorphism.

(3) Verify that the kernel of $F$ is $\{(1, 1), (-1, -1)\}$ and therefore that $F$ is two-to-one.

(4) How is $F$ related to the function $\varphi$ from the previous exercise?

*Comment: We will show later that the image of $F$ is $SO(4)$.*

**Ex. 3.18** (Gram-Schmidt)**.** For $m < n$, let $S = \{v_1, v_2, ..., v_m\} \subset \mathbb{K}^n$ be an orthonormal set.

(1) Prove that $S$ can be completed to an orthonormal basis; that is, there exist vectors $v_{m+1}, ..., v_n$ such that $\{v_1, ..., v_n\}$ is an orthonormal basis of $\mathbb{K}^n$.

(2) Prove that the vectors in part (1) can be chosen such that the matrix, $M \in \mathcal{O}_n(\mathbb{K})$, whose rows are $\{v_1, ..., v_n\}$ (in that order) has determinant 1.

(3) Re-phrase the case $m = 1$ of part (2) as follows: For any unit-length vector $v \in \mathbb{K}^n$, there exists a matrix $M \in \mathcal{O}_n(\mathbb{K})$ with determinant 1 such that $R_M(e_1) = v$.

Chapter 4

# The topology of matrix groups

This text is about the subgroups of $GL_n(\mathbb{K})$. So far, we have considered such a subgroup, $G$, as a purely algebraic object. Geometric intuition has been relevant only because $R_A$ is a *motion* of $\mathbb{K}^n$ for every $A \in G$.

We now begin to study $G$ as a geometric object. Since

$$G \subset GL_n(\mathbb{K}) \subset M_n(\mathbb{K}) \cong \mathbb{K}^{n^2} \cong \begin{cases} \mathbb{R}^{n^2} & \text{if } \mathbb{K} = \mathbb{R} \\ \mathbb{R}^{2n^2} & \text{if } \mathbb{K} = \mathbb{C} \\ \mathbb{R}^{4n^2} & \text{if } \mathbb{K} = \mathbb{H} \end{cases},$$

we can think of $G$ as a subset of a **Euclidean space**, meaning $\mathbb{R}^m$ for some $m$. Many familiar subsets of Euclidean spaces, like the sphere $S^n \subset \mathbb{R}^{n+1}$, or the graphs of functions of several variables, have visualizable shapes. It makes sense to ask "what is the shape of $G$?" For example, we previously recognized the shape of $Sp(1)$ as the three-dimensional sphere $S^3$.

In this chapter, we review some topology, which provides an ideal vocabulary for discussing the shape of a subset $G \subset \mathbb{R}^m$. Is it compact? path-connected? open? closed? We will define and briefly discuss these terms and apply them to subgroups of the general linear groups.

## 1. Open and closed sets and limit points

The natural **distance function** on $\mathbb{R}^m$ was defined in Section 5 of Chapter 3 as $\text{dist}(X,Y) = |X - Y|$. Its most important property is:

**Proposition 4.1** (The Triangle Inequality)**.** *For all $X, Y, Z \in \mathbb{R}^m$,*

$$dist(X,Z) \leq dist(X,Y) + dist(Y,Z).$$

**Proof.** For all $V, W \in \mathbb{R}^m$, the Schwarz inequality (Proposition 3.7) gives:

$$\begin{aligned} |V+W|^2 &= |V|^2 + 2\langle V, W\rangle + |W|^2 \\ &\leq |V|^2 + 2|V|\cdot|W| + |W|^2 = (|V|+|W|)^2. \end{aligned}$$

Thus, $|V+W| \leq |V| + |W|$. Applying this inequality to the vectors pictured in Figure 1 proves the triangle inequality. □

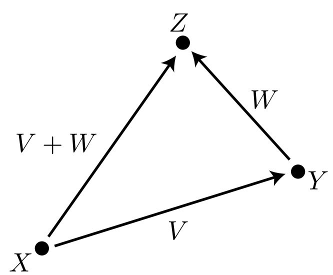



**Figure 1.** Proof of the triangle inequality.

Our study of topology begins with precise language for discussing whether a subset of Euclidean space contains its boundary points. First, for $p \in \mathbb{R}^m$ and $r > 0$, we denote the **ball** about $p$ of radius $r$ as:

$$B(p,r) = \{q \in \mathbb{R}^m \mid \text{dist}(p,q) < r\}.$$

In other words, $B(p,r)$ contains all points closer than a distance $r$ from $p$.

**Definition 4.2.** *A point $p \in \mathbb{R}^m$ is called a* ***boundary point*** *of a subset $S \subset \mathbb{R}^m$ if for all $r > 0$, the ball $B(p,r)$ contains at least one point in $S$ and at least one point not in $S$. The collection of all boundary points of $S$ is called the* ***boundary*** *of $S$.*

Sometimes, but not always, boundary points of $S$ are contained in $S$. For example, consider the "open upper half-plane"

$$H = \{(x, y) \in \mathbb{R}^2 \mid y > 0\},$$

and the "closed upper half-plane"

$$\overline{H} = \{(x, y) \in \mathbb{R}^2 \mid y \geq 0\}.$$

The $x$-axis, $\{(x, 0) \in \mathbb{R}^2\}$, is the boundary of $H$ and also of $\overline{H}$. So $H$ contains none of its boundary points, while $\overline{H}$ contains all of its boundary points. This distinction is so central, we introduce vocabulary for it:

**Definition 4.3.** *Let $S \subset \mathbb{R}^m$ be a subset.*

(1) *$S$ is called **open** if it contains none of its boundary points.*

(2) *$S$ is called **closed** if it contains all of its boundary points.*

In the previous example, $H$ is open, while $\overline{H}$ is closed. If part of the $x$-axis is adjoined to $H$ (say the positive part), the result is neither closed nor open, since it contains some of its boundary points but not all of them.

A set $S \subset \mathbb{R}^m$ and its complement $S^c = \{p \in \mathbb{R}^m \mid p \notin S\}$ clearly have the same boundary. If $S$ contains none of these common boundary points, then $S^c$ must contain all of them, and vice-versa. So we learn that:

**Proposition 4.4.** *A set $S \subset \mathbb{R}^m$ is closed if and only if its complement, $S^c$, is open.*

The following provides a useful alternative definition of "open":

**Proposition 4.5.** *A set $S \subset \mathbb{R}^m$ is open if and only if for all $p \in S$, there exists $r > 0$ such that $B(p, r) \subset S$.*

**Proof.** If $S$ is not open, then it contains at least one of its boundary points, and no ball about such a boundary point is contained in $S$. Conversely, suppose that there is a point $p \in S$ such that no ball about $p$ is contained in $S$. Then $p$ is a boundary point of $S$, so $S$ is not open. $\square$

The proposition says that if you live in an open set, then so do all of your sufficiently close neighbors. How close is sufficient depends on how close you live from the boundary. For example, the set

$$S = (0, \infty) \subset \mathbb{R}$$

is open because for any $x \in S$, the ball $B(x, x/2) = (x/2, 3x/2)$ lies inside of $S$. When $x$ is close to 0, the radius of this ball is small.

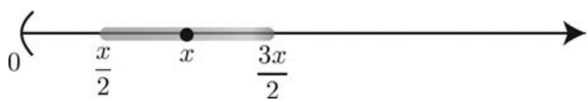



**Figure 2.** The set $(0, \infty) \subset \mathbb{R}$ is open because it contains a ball about each of its points.

Similarly, for any $p \in \mathbb{R}^m$ and any $r > 0$, the ball $B = B(p, r)$ is itself open because about any $q \in B$, the ball of radius $(r - \text{dist}(p, q))/2$ lies in $B$ (by the triangle inequality).

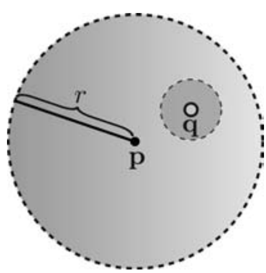



**Figure 3.** The set $B(p, r) \subset \mathbb{R}^m$ is open because it contains a ball about each of its points.

The collection of all open subsets of $\mathbb{R}^m$ is called the **topology** of $\mathbb{R}^m$. It is surprising how many important concepts are topological, that is, definable purely in terms of the topology of $\mathbb{R}^m$. For example, the notion of whether a subset is closed is topological. The distance between points of $\mathbb{R}^m$ is not topological. The notion of *convergence* is topological by the second definition below, although it may not initially seem so from the first:

**Definition 4.6.** *An infinite sequence $\{p_1, p_2, ...\}$ of points of $\mathbb{R}^m$ is said to **converge** to $p \in \mathbb{R}^m$ if either of the following equivalent conditions hold:*

(1) $\lim_{n\to\infty} dist(p, p_n) = 0$.

(2) *For every open set, $U$, containing $p$, there exists an integer $N$ such that $p_n \in U$ for all $n > N$.*

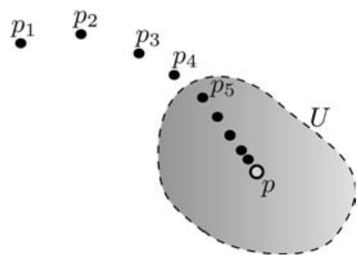



**Figure 4.** A convergent sequence is eventually inside of any open set containing its limit.

**Definition 4.7.** *A point $p \in \mathbb{R}^m$ is called a **limit point** of a subset $S \subset \mathbb{R}^m$ if there exists an infinite sequence of points of $S$ that converges to $p$.*

Any point $p \in S$ is a limit point of $S$, as evidenced by the redundant infinite sequence $\{p, p, p, ...\}$. Any point of the boundary of $S$ is a limit point of $S$ as well (why?). In fact, the collection of limit points of $S$ equals the union of $S$ and the boundary of $S$. Therefore, *a set $S \subset \mathbb{R}^m$ is closed if and only if it contains all of its limit points*, since this is the same as requiring it to contain all of its boundary points.

It is possible to show that a sequence converges without knowing its limit just by showing that the terms get closer and closer to each other; more precisely,

**Definition 4.8.** *An infinite sequence of points $\{p_1, p_2, ...\}$ in $\mathbb{R}^m$ is called a **Cauchy sequence** if for every $\epsilon > 0$ there exists an integer $N$ such that $dist(p_i, p_j) < \epsilon$ for all $i, j > N$.*

It is straightforward to prove that any convergent sequence is Cauchy. A fundamental property of Euclidean space is the converse:

**Proposition 4.9.** *Any Cauchy sequence in $\mathbb{R}^m$ converges to some point of $\mathbb{R}^m$.*

We end this section with an important *relative* notion of open and closed:

**Definition 4.10.** *Let $S \subset G \subset \mathbb{R}^m$ be subsets.*

(1) *$S$ is called **open in** $G$ if there exists an open subset of $\mathbb{R}^m$ whose intersection with $G$ equals $S$.*
(2) *$S$ is called **closed in** $G$ if there exists a closed subset of $\mathbb{R}^m$ whose intersection with $G$ equals $S$.*

For example, the hemisphere $\{(x, y, z) \in S^2 \subset \mathbb{R}^3 \mid z > 0\}$ is open in $S^2$ because it is the intersection with $S^2$ of the open set $\{(x, y, z) \in \mathbb{R}^3 \mid z > 0\}$.

It is straightforward to show that $S$ is open in $G$ if and only if its *relative complement* $\{p \in G \mid p \notin S\}$ is closed in $G$. The following is a useful equivalent characterization of "open in" and "closed in":

**Proposition 4.11.** *Let $S \subset G \subset \mathbb{R}^m$ be subsets.*

(1) *$S$ is open in $G$ if and only if for all $p \in S$, there exists $r > 0$ such that $\{q \in G \mid dist(p, q) < r\} \subset S$.*
(2) *$S$ is closed in $G$ if and only if every $p \in G$ that is a limit point of $S$ is contained in $S$.*

Part (1) says that if you live in a set that's open in $G$, then so do all of your sufficiently close neighbors in $G$. Part (2) says that if you live in a set that's closed in $G$, then you contain all of your limit points that belong to $G$. For example, the interval $[0, 1)$ is neither open nor closed in $\mathbb{R}$, but is open in $[0, 2]$ and is closed in $(-1, 1)$.

A set $S$ is called **dense** in $G$ if every point of $G$ is a limit point of $S$. For example, the irrational numbers are dense in $\mathbb{R}$.

Let $p \in G \subset \mathbb{R}^m$. A **neighborhood** of $p$ in $G$ means a subset of $G$ that is open in $G$ and contains $p$. For example, $(1 - \epsilon, 1 + \epsilon)$

is a neighborhood of 1 in $(0, 2)$ for any $\epsilon \in (0, 1]$. Also, $[0, \epsilon)$ is a neighborhood of 0 in $[0, 1]$ for any $\epsilon \in (0, 1]$.

The collection of all subsets of $G$ that are open in $G$ is called the **topology** of $G$. In the remainder of this chapter, pay attention to which properties of a set $G$ are topological, that is, definable in terms of only the topology of $G$. For example, the notion of a sequence of points of $G$ converging to $p \in G$ is topological. Why? Because convergence means that the sequence is eventually inside of any neighborhood of $p$ in $\mathbb{R}^m$; this is the same as being eventually inside of any neighborhood of $p$ in $G$, which has only to do with the topology of $G$. The idea is to forget about the ambient $\mathbb{R}^m$ and regard $G$ as an independent object with a topology and hence a notion of convergence.

## 2. Continuity

Let $G_1 \subset \mathbb{R}^{m_1}$ and $G_2 \subset \mathbb{R}^{m_2}$. A function $f : G_1 \to G_2$ is called *continuous* if it maps nearby points to nearby points; more precisely:

**Definition 4.12.** *A function $f : G_1 \to G_2$ is called **continuous** if for any infinite sequence $\{p_1, p_2, ...\}$ of points in $G_1$ that converges to a point $p \in G_1$, the sequence $\{f(p_1), f(p_2), ...\}$ converges to $f(p)$.*

For example, the "step function" $f : \mathbb{R} \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

is not continuous. Why? Because the sequence

$$\{1/2, 1/3, 1/4, ...\}$$

in the domain of $f$ converges to 0, but the images

$$\{f(1/2) = 1, f(1/3) = 1, f(1/4) = 1, ...\}$$

converge to 1 rather than to $f(0) = 0$.

Notice that $f$ is continuous if and only if it is continuous when regarded as a function from $G_1$ to $\mathbb{R}^{m_2}$. It is nevertheless useful to forget about the ambient Euclidean spaces and regard $G_1$ and $G_2$ as independent objects. This vantage point leads to the following beautiful, although less intuitive, way to define continuity:

**Proposition 4.13.** *For $f : G_1 \to G_2$, the following are equivalent:*

(1) *$f$ is continuous.*

(2) *For any set $U$ that's open in $G_2$, $f^{-1}(U)$ is open in $G_1$.*

(3) *For any set $U$ that's closed in $G_2$, $f^{-1}(U)$ is closed in $G_1$.*

Here, $f^{-1}(U)$ denotes the set $\{p \in G_1 \mid f(p) \in U\}$. The above step function fails this continuity test because

$$f^{-1}((-1/2, 1/2)) = (-\infty, 0],$$

which is not open.

It is now clear that continuity is a topological concept, since this alternative definition involved only the topologies of $G_1$ and $G_2$.

Familiar functions from $\mathbb{R}$ to $\mathbb{R}$, like polynomial, rational, trigonometric, exponential, and logarithmic functions, are all continuous on their domains. It is straightforward to prove that:

**Proposition 4.14.** *The composition of two continuous functions is continuous.*

We next wish to describe what it means for $G_1$ and $G_2$ to be "topologically the same". There should be a bijection between them that pairs open sets with open sets. More precisely,

**Definition 4.15.** *A function $f : G_1 \to G_2$ is called a* ***homeomorphism*** *if $f$ is bijective and continuous and $f^{-1}$ is continuous. If such a function exists, then $G_1$ and $G_2$ are said to be* ***homeomorphic****.*

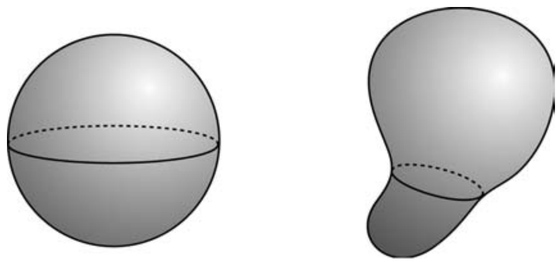


**Figure 5.** Homeomorphic sets.

Homeomorphic sets have the same "essential shape," like the two subsets of $\mathbb{R}^3$ pictured in Figure 5. The hypothesis that $f^{-1}$ be continuous is necessary. To see this, consider the function $f : [0, 2\pi) \to S^1 \subset \mathbb{R}^2$ defined as $f(t) = (\cos t, \sin t)$. It is straightforward to check that $f$ is continuous and bijective, but $f^{-1}$ is not continuous (why not?). We will see in Section 4 that $[0, 2\pi)$ is not homeomorphic to $S^1$, since only the latter is compact.

## 3. Path-connected sets

**Definition 4.16.** *A subset $G \subset \mathbb{R}^m$ is called **path-connected** if for every pair $p, q \in G$, there exists a continuous function $f : [0,1] \to G$ with $f(0) = p$ and $f(1) = q$.*

The terminology comes from visualizing the image of such an $f$ as a "path" in $G$ beginning at $p$ and ending at $q$.

For example, the disk $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$ is path-connected, since any pair $p, q \in A$ can be connected by the straight line segment between them, explicitly parameterized as

$$f(t) = p + t(q - p).$$

But the disjoint union of two discs,

$$B = \{p \in \mathbb{R}^2 \mid \text{dist}(p, (-2, 0)) < 1 \text{ or } \text{dist}(p, (2, 0)) < 1\},$$

is not path-connected, because no continuous path exists between points in different disks (why not?).

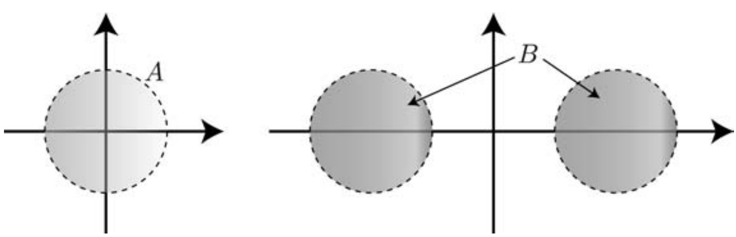



**Figure 6.** $A$ is path-connected, while $B$ is not.

In the non-path-connected example, the right disk is **clopen** (both open and closed) in $B$, and therefore so too is the left disk.

In other words, $B$ decomposes into the disjoint union of two subsets that are both clopen in $B$. Such a separation of a path-connected set is impossible:

**Proposition 4.17.** *A path-connected set $G \subset \mathbb{R}^m$ has no clopen subsets other than itself and the empty set.*

**Proof.** We first prove that the interval $[0,1]$ has no clopen subsets other than itself and the empty set. Suppose $A \subset [0,1]$ is another one. Let $t$ denote the infimum of $A$. Since $A$ is closed, $t \in A$. Since $A$ is open, there exists $r > 0$ such that all points of $[0,1]$ with distance $< r$ from $t$ lie in $A$. This contradicts the fact that $t$ is the infimum of $A$ unless $t = 0$. Therefore, $0 \in A$. Since the complement $A^c$ of $A$ is also clopen, the same argument proves that $0 \in A^c$, which is impossible.

Next, let $G \subset \mathbb{R}^m$ be any path-connected set. Suppose that $A \subset G$ is a clopen subset. Suppose there exist points $p, q \in G$ such that $p \in A$ and $q \notin A$. Since $G$ is path-connected, there exists a continuous function $f : [0,1] \to G$ with $f(0) = p$ and $f(1) = q$. Then $f^{-1}(A)$ is a clopen subset of $[0,1]$ that contains 0 but not 1, contradicting the previous paragraph. □

In practice, to prove that a property is true at all points in a path-connected set, it is often convenient to prove that the set of points where the property holds is non-empty, open, and closed.

Since continuity is a topological notion, so is path-connectedness. In particular,

**Proposition 4.18.** *If $G_1 \subset \mathbb{R}^{m_1}$ and $G_2 \subset \mathbb{R}^{m_2}$ are homeomorphic, then either both are path-connected or neither is path-connected.*

## 4. Compact sets

The notion of compactness is fundamental to topology. We begin with the most intuitive definition.

**Definition 4.19.** *A subset $G \subset \mathbb{R}^m$ is called **bounded** if $G \subset B(p,r)$ for some $p \in \mathbb{R}^m$ and some $r > 0$. Further, $G$ is called **compact** if it is closed and bounded.*

Compact sets are those that contain their limit points and lie in a finite chunk of Euclidean space. Unfortunately, this definition is not topological, since "bounded" cannot be defined without referring to the distance function on $\mathbb{R}^m$. In particular, boundedness is not preserved by homeomorphisms, since the bounded set $(0,1)$ is homeomorphic to the unbounded set $\mathbb{R}$. Nevertheless, compactness is a topological notion, as is shown by the following alternative definition:

**Definition 4.20.** *Let $G \subset \mathbb{R}^m$.*

(1) *An **open cover** of $G$ is a collection, $\mathbb{O}$, of sets that are open in $G$, whose union equals $G$.*

(2) *$G$ is called **compact** if every open cover, $\mathbb{O}$, of $G$ has a finite subcover, meaning a finite subcollection $\{U_1, ..., U_n\} \subset \mathbb{O}$ whose union equals $G$.*

The equivalence of our two definitions of compactness is called the **Heine-Borel Theorem**. The easy half of its proof goes like this: Suppose that $G$ is not bounded. Then the collection

$$\{p \in G \mid \text{dist}(0,p) < n\},$$

for $n = 1, 2, 3, ...$, is an open cover of $G$ with no finite subcover. Next suppose that $G$ is not closed, which means it is missing a limit point $q \in \mathbb{R}^m$. Then the collection $\{p \in G \mid \text{dist}(p,q) > 1/n\}$, for $n = 1, 2, 3, ...$, is an open cover of $G$ with no finite subcover.

The other half of the proof is substantially more difficult. We content ourselves with a few examples.

The open interval $(0,1) \subset \mathbb{R}$ is not compact because it is not closed, or alternatively because

$$\mathbb{O} = \{(0,1/2), (0,2/3), (0,3/4), (0,4/5)...\}$$

is an open cover of $(0,1)$ that has no finite subcover.

The closed interval $[0,1]$ is compact because it is closed and bounded. It is somewhat difficult to prove directly that every open cover of $[0,1]$ has a finite subcover; attempting to do so will increase your appreciation of the Heine-Borel Theorem.

Since our second definition of compactness is topological, it is straightforward to prove that:

**Proposition 4.21.** *If $G_1 \subset \mathbb{R}^{m_1}$ and $G_2 \subset \mathbb{R}^{m_2}$ are homeomorphic, then either both are compact or neither is compact.*

There is a third useful characterization of compactness, which depends on the notion of *subconvergence*.

**Definition 4.22.** *An infinite sequence of points $\{p_1, p_2, p_3, ...\}$ in $\mathbb{R}^m$ is said to* ***subconverge*** *to $p \in \mathbb{R}^m$ if there is an infinite subsequence, $\{p_{i_1}, p_{i_2}, p_{i_3}, ...\}$ (with $i_1 < i_2 < i_3 < \cdots$) that converges to $p$.*

**Proposition 4.23.** *A subset $G \subset \mathbb{R}^m$ is compact if and only if every infinite sequence of points in $G$ subconverges to some $p \in G$.*

For example, the sequence $\{1/2, 2/3, 3/4, ...\}$ in $G = (0, 1)$ subconverges only to $1 \notin G$, which gives another proof that $(0, 1)$ is not compact.

The next proposition says that the continuous image of a compact set is compact.

**Proposition 4.24.** *Let $G \subset \mathbb{R}^{m_1}$. Let $f : G \to \mathbb{R}^{m_2}$ be continuous. If $G$ is compact, then the image $f(G)$ is compact.*

**Proof.** The function $f$ is also continuous when regarded as a function from $G$ to $f(G)$. Let $\mathbb{O}$ be an open cover of $f(G)$. Then $f^{-1}(U)$ is open in $G$ for every $U \in \mathbb{O}$, so $f^{-1}(\mathbb{O}) = \{f^{-1}(U) \mid U \in \mathbb{O}\}$ is an open cover of $G$. Since $G$ is compact, there exists a finite subcover $\{f^{-1}(U_1), ..., f^{-1}(U_n)\}$ of $f^{-1}(\mathbb{O})$. It is straightforward to check that $\{U_1, U_2, ..., U_n\}$ is a finite subcover of $\mathbb{O}$. □

**Corollary 4.25.** *If $G \subset \mathbb{R}^m$ is compact and $f : G \to \mathbb{R}$ is continuous, then $f$ attains is supremum and infimum.*

The conclusion that $f$ attains is supremum means two things. First, the supremum of $f(G)$ is finite (because $f(G)$ is bounded). Second, there is a point $p \in G$ for which $f(p)$ equals this supremum (because $f(G)$ is closed).

## 5. Definition and examples of matrix groups

As mentioned earlier in this chapter, a subgroup $G \subset GL_n(\mathbb{K})$ can be considered a subset of Euclidean space, so we can ask whether it is open, closed, path-connected, compact, etc. The title of this book comes from:

**Definition 4.26.** *A* ***matrix group*** *is a subgroup $G \subset GL_n(\mathbb{K})$ that is closed in $GL_n(\mathbb{K})$.*

The "closed" hypothesis means that if a sequence of matrices in $G$ has a limit in $GL_n(\mathbb{K})$, then that limit must lie in $G$. In other words, $G$ contains all of its non-singular limit points.

We now verify that several previously introduced subgroups of $GL_n(\mathbb{K})$ are closed and are therefore matrix groups.

**Proposition 4.27.** *$\mathcal{O}_n(\mathbb{K})$, $SL_n(\mathbb{K})$, $SO(n)$ and $SU(n)$ are matrix groups.*

**Proof.** We must prove that each is closed in $GL_n(\mathbb{K})$. For $\mathcal{O}_n(\mathbb{K})$, define $f : M_n(\mathbb{K}) \to M_n(\mathbb{K})$ as $f(A) = A \cdot A^*$. This function $f$ is continuous, because for each $i, j$, the $\mathbb{K}$-valued function

$$f_{ij}(A) = (A \cdot A^*)_{ij}$$

is continuous because it is a polynomial in the entries of $A$. The single-element set $\{I\} \subset M_n(\mathbb{K})$ is closed, so $\mathcal{O}_n(\mathbb{K}) = f^{-1}(\{I\})$ is closed in $M_n(\mathbb{K})$ and is therefore closed in $GL_n(\mathbb{K})$.

For $SL_n(\mathbb{K})$, we first prove the function $\det : M_n(\mathbb{K}) \to \mathbb{R}$ or $\mathbb{C}$ is continuous. When $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, this is because $\det(A)$ is an n-degree polynomial in the entries of $A$ by Equation 1.5. When $\mathbb{K} = \mathbb{H}$, this is because $\det(A)$ is shorthand for $\det(\Psi_n(A))$, and the composition of two continuous functions is continuous. Since the single-element set $\{1\}$ is closed, $SL_n(\mathbb{K}) = \det^{-1}(\{1\})$ is closed in $M_n(\mathbb{K})$ and therefore also in $GL_n(\mathbb{K})$.

For $SO(n)$ and $SU(n)$, notice that $SO(n) = O(n) \cap SL_n(\mathbb{R})$ and $SU(n) = U(n) \cap SL_n(\mathbb{C})$, and the intersection of two closed sets is closed. $\square$

In the remainder of this book, we will emphasize compact matrix groups, so the following proposition is crucial:

**Proposition 4.28.** *Each of the groups $O(n)$, $SO(n)$, $U(n)$, $SU(n)$ and $Sp(n)$ is compact for any $n$.*

**Proof.** In proving above that these groups are closed in $GL_n(\mathbb{K})$, we actually proved the stronger fact that they are closed in the Euclidean space $M_n(\mathbb{K})$. So it remains to prove that these groups are bounded, which follows from the fact that each row of $A \in \mathcal{O}_n(\mathbb{K})$ is unit-length, according to part (3) of Proposition 3.9. □

In the exercises, you will verify that several other familiar matrix groups are non-compact, like $GL_n(\mathbb{K})$ for $n \geq 1$ and $SL_n(\mathbb{K})$ for $n \geq 2$.

Why did we define matrix groups to be closed in $GL_n(\mathbb{K})$? Because, as we will see later, non-closed subgroups are not necessarily manifolds. Exercises 4.23 and 4.24 exhibit the bad behavior of non-closed subgroups that underlies this fact. Nevertheless, the hypothesis that matrix groups are closed will not be used until Chapter 7. Until then, the facts we prove about matrix groups will also be true for non-closed subgroups of $GL_n(\mathbb{K})$.

## 6. Exercises

**Ex. 4.1.** Prove Proposition 4.11.

**Ex. 4.2.** Let $S \subset G \subset \mathbb{R}^m$. Prove that $S$ is open in $G$ if and only if $\{p \in G \mid p \notin S\}$ is closed in $G$.

**Ex. 4.3.** Prove Proposition 4.13.

**Ex. 4.4.** Prove Proposition 4.14.

**Ex. 4.5.** Prove Proposition 4.18.

**Ex. 4.6.** Prove Proposition 4.21.

**Ex. 4.7.** Prove that $GL_n(\mathbb{K})$ is open in $M_n(\mathbb{K})$.

**Ex. 4.8.** Prove that $GL_n(\mathbb{K})$ is non-compact when $n \geq 1$. Prove that $SL_n(\mathbb{K})$ is non-compact when $n \geq 2$. What about $SL_1(\mathbb{K})$?

**Ex. 4.9.** Let $G$ be a matrix group. Prove that a subgroup $H \subset G$ that is closed in $G$ is itself a matrix group.

**Ex. 4.10.** Prove that $SO(n)$ and

$$O(n)^- = \{A \in O(n) \mid \det(A) = -1\}$$

are both clopen in $O(n)$.

**Ex. 4.11.** Prove that $\mathrm{Aff}_n(\mathbb{K}) \subset GL_{n+1}(\mathbb{K})$ (defined in Exercise 3.11) is a matrix group. Show that $\mathrm{Aff}_n(\mathbb{K})$ is NOT closed in $M_{n+1}(\mathbb{K})$. Is $\mathrm{Aff}_n(\mathbb{K})$ compact?

**Ex. 4.12.** A matrix $A \in M_n(\mathbb{K})$ is called **upper triangular** if all entries below the diagonal are zero; i.e., $A_{ij} = 0$ for all $i > j$. Prove that the following is a matrix group:

$$UT_n(\mathbb{K}) = \{A \in GL_n(\mathbb{K}) \mid A \text{ is upper triangular}\}.$$

Show that $UT_n(\mathbb{K})$ is not closed in $M_n(\mathbb{K})$. Is $UT_n(\mathbb{K})$ compact?

**Ex. 4.13.** Prove that $\mathrm{Isom}(\mathbb{R}^n)$ is a matrix group. Is it compact?

**Ex. 4.14.** Prove that $SO(3)$ is path-connected.

**Ex. 4.15.** Prove that $Sp(1)$ is path-connected.

**Ex. 4.16.** Prove that the image under a continuous function of a path-connected set is path-connected.

**Ex. 4.17.** We will prove later that $Sp(n)$ is path-connected. Assuming this, and using Propositions 2.10 and 3.12, prove that the determinant of any $A \in Sp(n)$ equals 1.

**Ex. 4.18.** Prove that $\mathcal{O}_n(\mathbb{K})$ is isomorphic to a subgroup of $\mathcal{O}_{n+1}(\mathbb{K})$.

**Ex. 4.19.** Prove that $U(n)$ is isomorphic to a subgroup of $SU(n+1)$.

**Ex. 4.20.** Let $G \subset GL_n(\mathbb{R})$ be a compact subgroup.

(1) Prove that every element of $G$ has determinant 1 or $-1$.

(2) Must it be true that $G \subset O(n)$?
*Hint: Consider conjugates of $O(n)$.*

**Ex. 4.21.** There are two natural functions from $SU(n) \times U(1)$ to $U(n)$. The first is $f_1(A, (\lambda)) = \lambda \cdot A$. The second is $f_2(A, (\lambda)) =$ the result of multiplying each entry of the first row of $A$ times $\lambda$.

(1) Prove that $f_1$ is an $n$-to-1 homomorphism.

(2) Prove that $f_2$ is a homeomorphism, but is not a homomorphism when $n \geq 2$.

*Comment: Later we will prove that $U(n)$ is not isomorphic to $SU(n) \times U(1)$, even though they are homeomorphic.*

**Ex. 4.22.** $SO(2)$ is a subgroup of $SL_2(\mathbb{R})$. Another is:

$$H = \left\{ \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \in M_2(\mathbb{R}) \middle| a > 0 \right\}.$$

Prove that the function $f : SO(2)\times H \to SL_2(\mathbb{R})$ defined as $f(A,B) = A \cdot B$ is a homeomorphism, but not a homomorphism.
*Comment: The **polar decomposition** theorem states that $SL_n(\mathbb{R})$ is homeomorphic to $SO(n)$ times a Euclidean space.*

**Ex. 4.23.** Let $\lambda \in \mathbb{R}$ be an irrational multiple of $2\pi$. Define

$$G = \{(e^{\lambda t \mathbf{i}}) | t \in \mathbb{Z}\} \subset U(1) \subset GL_1(\mathbb{C}).$$

Prove that $G$ is a subgroup of $GL_1(\mathbb{C})$, but not a matrix group. Prove that $G$ is dense in $U(1)$.

**Ex. 4.24.** Let $\lambda \in \mathbb{R}$ be an irrational number. Define

$$G = \left\{ \begin{pmatrix} e^{t\mathbf{i}} & 0 \\ 0 & e^{\lambda t \mathbf{i}} \end{pmatrix} \middle| t \in \mathbb{R} \right\} \subset \overline{G} = \left\{ \begin{pmatrix} e^{t\mathbf{i}} & 0 \\ 0 & e^{s\mathbf{i}} \end{pmatrix} \middle| t, s \in \mathbb{R} \right\} \subset GL_2(\mathbb{C}).$$

(1) Prove that $G$ and $\overline{G}$ are subgroups of $GL_2(\mathbb{C})$.
(2) Prove that $G$ is dense in $\overline{G}$.
(3) Define $f : \mathbb{R} \to G$ as follows: $f(t) = \begin{pmatrix} e^{t\mathbf{i}} & 0 \\ 0 & e^{\lambda t \mathbf{i}} \end{pmatrix}$. Show that $f$ is an isomorphism (with $\mathbb{R}$ considered a group under addition), but not a homeomorphism.

**Ex. 4.25.** Let $G \subset GL_n(\mathbb{R})$ denote the set of matrices whose determinants are integer powers of 2. Is $G$ a matrix group?

**Ex. 4.26.** Prove or find a counterexample of each statement:

(1) If $X \subset \mathbb{R}^n$ is compact, then $\mathrm{Symm}(X)$ is compact.
(2) If $\mathrm{Symm}(X)$ is compact, then $X$ is compact.

**Ex. 4.27.** Prove that the image of the function $\varphi : Sp(1) \to O(3)$ defined in Exercise 3.16 is contained in $SO(3)$.
*Hint: Use Exercise* 4.15.

**Ex. 4.28.** Prove that the image of the function $F : Sp(1) \times Sp(1) \to O(4)$ defined in Exercise 3.17 is contained in $SO(4)$.

# Chapter 5

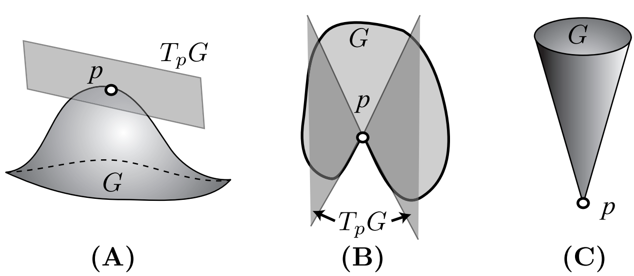
# Lie algebras

A matrix group $G \subset GL_n(\mathbb{K})$ is a subset of the Euclidean space $M_n(\mathbb{K})$, so we can discuss its tangent spaces.

**Definition 5.1.** *Let $G \subset \mathbb{R}^m$ be a subset, and let $p \in G$. The **tangent space** to $G$ at $p$ is:*

$$T_pG = \{\gamma'(0) \mid \gamma : (-\epsilon, \epsilon) \to G \text{ is differentiable with } \gamma(0) = p\}.$$

In other words, $T_pG$ means the set of initial velocity vectors of differentiable paths though $p$ in $G$. The term *differentiable* means that, when we consider $\gamma$ as a path in $\mathbb{R}^m$, the $m$ components of $\gamma$ are differentiable functions from $(-\epsilon, \epsilon)$ to $\mathbb{R}$.




If $G \subset \mathbb{R}^3$ is the graph of a differentiable function of two variables, then $T_pG$ is a 2-dimensional subspace of $\mathbb{R}^3$, as in Figure A (subspaces always pass though the origin; its translate to $p$ is actually what is drawn). For a general subset $G \subset \mathbb{R}^m$, $T_pG$ is not necessarily a subspace of $\mathbb{R}^m$. In Figure B, the tangent space is two sectors, while in Figure C, the tangent space is $\{0\}$.

**Definition 5.2.** *The **Lie algebra** of a matrix group $G \subset GL_n(\mathbb{K})$ is the tangent space to $G$ at $I$. It is denoted $\mathfrak{g} = \mathfrak{g}(G) = T_IG$.*

In this chapter, we prove that $\mathfrak{g}$ is a subspace of the Euclidean space $M_n(\mathbb{K})$. This is our first evidence that matrix groups are "nice" sets (you should picture them like Figure A, not like B or C; we will make this precise when we prove that matrix groups are manifolds in Chapter 7). We also describe the Lie algebras of many familiar matrix groups.

The Lie algebra is an indispensable tool for studying a matrix group. It contains a surprising amount of information about the group, especially together with the Lie bracket operation, which we will discuss in Chapter 8. In much of the remainder of this book, we will learn about matrix groups by studying their Lie algebras.

## 1. The Lie algebra is a subspace

Let $G \subset GL_n(\mathbb{K}) \subset M_n(\mathbb{K})$ be a matrix group. At the beginning of Chapter 4, we described how $M_n(\mathbb{K})$ can be identified with a Euclidean space. For example, $M_2(\mathbb{C}) \cong \mathbb{R}^8$ via the identification:

$$\begin{pmatrix} a+b\mathbf{i} & c+d\mathbf{i} \\ e+f\mathbf{i} & g+h\mathbf{i} \end{pmatrix} \leftrightarrow (a,b,c,d,e,f,g,h).$$

This identification allows us to talk about tangent vectors to differentiable paths in $M_n(\mathbb{K})$. For example a differentiable path in $M_2(\mathbb{C})$ has the form:

$$\gamma(t) = \begin{pmatrix} a(t)+b(t)\mathbf{i} & c(t)+d(t)\mathbf{i} \\ e(t)+f(t)\mathbf{i} & g(t)+h(t)\mathbf{i} \end{pmatrix},$$

where $a(t)$ through $h(t)$ are differentiable functions. The derivative is:

$$\gamma'(t) = \begin{pmatrix} a'(t)+b'(t)\mathbf{i} & c'(t)+d'(t)\mathbf{i} \\ e'(t)+f'(t)\mathbf{i} & g'(t)+h'(t)\mathbf{i} \end{pmatrix}.$$

Matrix multiplication interacts with differentiation in the following way.

**Proposition 5.3** (The product rule)**.** *If $\gamma, \beta : (-\epsilon, \epsilon) \to M_n(\mathbb{K})$ are differentiable, then so is the product path $(\gamma \cdot \beta)(t) = \gamma(t) \cdot \beta(t)$, and*

$$(\gamma \cdot \beta)'(t) = \gamma(t) \cdot \beta'(t) + \gamma'(t) \cdot \beta(t).$$

**Proof.** When $n = 1$ and $\mathbb{K} = \mathbb{R}$, this is the familiar product rule from calculus. When $n = 1$ and $\mathbb{K} = \mathbb{C}$, we denote $\gamma(t) = a(t) + b(t)\mathbf{i}$ and $\beta(t) = c(t) + d(t)\mathbf{i}$. Omitting the $t$'s to shorten notation, we have:

$$\begin{aligned}
(\gamma \cdot \beta)' &= ((ac - bd) + (ad + bc)\mathbf{i})' \\
&= (ac' + a'c - bd' - b'd) + (ad' + a'd + bc' + b'c)\mathbf{i} \\
&= ((ac' - bd') + (ad' + bc')\mathbf{i}) + ((a'c - b'd) + (a'd + b'c))\,\mathbf{i} \\
&= \gamma \cdot \beta' + \gamma' \cdot \beta.
\end{aligned}$$

When $n = 1$ and $\mathbb{K} = \mathbb{H}$, an analogous argument works. This completes the $n = 1$ case. For the general case, since

$$((\gamma \cdot \beta)(t))_{ij} = \sum_{l=1}^{n} \gamma(t)_{il} \cdot \beta(t)_{lj},$$

the derivative is:

$$\begin{aligned}
((\gamma \cdot \beta)'(t))_{ij} &= \sum_{l=1}^{n} \gamma(t)_{il} \cdot \beta'(t)_{lj} + \gamma'(t)_{il} \cdot \beta(t)_{lj} \\
&= (\gamma(t) \cdot \beta'(t))_{ij} + (\gamma'(t) \cdot \beta(t))_{ij}.
\end{aligned}$$

□

If $\gamma : (-\epsilon, \epsilon) \to GL_n(\mathbb{K})$ is a differentiable path, so is the inverse path $t \mapsto \gamma(t)^{-1}$ (see Exercise 5.16). The product rule gives:

$$0 = \frac{d}{dt}\left(\gamma(t)\gamma(t)^{-1}\right) = \gamma'(t)\gamma(t)^{-1} + \gamma(t)\frac{d}{dt}\left(\gamma(t)^{-1}\right).$$

When $\gamma(0) = I$, the solution is particularly clean:

$$\left.\frac{d}{dt}\right|_{t=0}\left(\gamma(t)^{-1}\right) = -\gamma'(0). \tag{5.1}$$

In other words, the inverse of a path through $I$ goes through $I$ in the opposite direction.

Another consequence of the product rule is the main result of this section:

**Proposition 5.4.** *The Lie algebra $\mathfrak{g}$ of a matrix group $G \subset GL_n(\mathbb{K})$ is a real subspace of $M_n(\mathbb{K})$.*

**Proof.** Let $\lambda \in \mathbb{R}$ and $A \in \mathfrak{g}$, which means that $A = \gamma'(0)$ for some differentiable path $\gamma(t)$ in $G$ with $\gamma(0) = I$. The path $\sigma(t) = \gamma(\lambda \cdot t)$ has initial velocity vector $\sigma'(0) = \lambda \cdot A$, which proves that $\lambda \cdot A \in \mathfrak{g}$.

Next let $A, B \in \mathfrak{g}$, which means that $A = \gamma'(0)$ and $B = \beta'(0)$ for some differentiable paths $\gamma, \beta$ in $G$ with $\gamma(0) = \beta(0) = I$. The product path $\sigma(t) = \gamma(t) \cdot \beta(t)$ is differentiable and lies in $G$. By the product rule, $\sigma'(0) = A + B$, which shows that $A + B \in \mathfrak{g}$. $\square$

The fact that Lie algebras are vector spaces over $\mathbb{R}$ allows us to define an important measurement of the size of a matrix group:

**Definition 5.5.** *The **dimension** of a matrix group $G$ means the dimension of its Lie algebra.*

Even though $M_n(\mathbb{C}) \cong \mathbb{C}^{n^2}$ is a vector space over $\mathbb{C}$ (rather than just a vector space over $\mathbb{R}$), the Lie algebra of a complex matrix group $G \subset GL_n(\mathbb{C})$ is NOT necessarily a $\mathbb{C}$-subspace of $M_n(\mathbb{C})$. Similarly, the Lie algebra of a quaternionic matrix group need not be an $\mathbb{H}$-subspace of $M_n(\mathbb{H})$. The dimension of a matrix group always means the dimension of its Lie algebra regarded as a REAL vector space.

## 2. Some examples of Lie algebras

In this section, we describe the Lie algebras of three familiar matrix groups. Lie algebras are denoted in lower case; for example, $gl_n(\mathbb{K})$ denotes the Lie algebra of $GL_n(\mathbb{K})$.

**Proposition 5.6.** $gl_n(\mathbb{K}) = M_n(\mathbb{K})$. *In particular,* $\dim(GL_n(\mathbb{R})) = n^2$, $\dim(GL_n(\mathbb{C})) = 2n^2$ *and* $\dim(GL_n(\mathbb{H})) = 4n^2$.

**Proof.** Let $A \in M_n(\mathbb{K})$. The path $\gamma(t) = I + t \cdot A$ in $M_n(\mathbb{K})$ satisfies $\gamma(0) = I$ and $\gamma'(0) = A$. Also, $\gamma$ restricted to a sufficiently small interval $(-\epsilon, \epsilon)$ lies in $GL_n(\mathbb{K})$. To justify this, notice $\det(\gamma(0)) = 1$. Since the determinant function is continuous, $\det(\gamma(t))$ is close to 1 (and is therefore non-zero) for $t$ close to 0. Thus, $A \in gl_n(\mathbb{K})$. $\square$

The general linear groups are large; all matrices are tangent to paths in them. But matrices in the Lie algebras of other matrix groups have special forms.

**Proposition 5.7.** *The Lie algebra, $u(1)$, of $U(1)$ equals span$\{(\mathbf{i})\}$, so* $\dim(U(1)) = 1$.

**Proof.** The path $\gamma(t) = (e^{\mathbf{i}t})$ in $U(1)$ satisfies $\gamma(0) = I$ and has $\gamma'(0) = (\mathbf{i})$, so $(\mathbf{i}) \in u(1)$. Therefore $\text{span}\{(\mathbf{i})\} \subset u(1)$. For the other inclusion, let $\gamma(t) = (a(t) + b(t)\mathbf{i})$ be a differentiable path in $U(1)$ with $\gamma(0) = I = (1)$. Since $|\gamma(t)|^2 = a(t)^2 + b(t)^2 = 1$, the value $a(0) = 1$ must be a local maximum of $a(t)$, so $a'(0) = 0$. Therefore $\gamma'(0) \in \text{span}\{(\mathbf{i})\}$. $\square$

A similar argument shows that $\dim(SO(2)) = 1$. We will see later that smoothly isomorphic matrix groups have the same dimension.

**Proposition 5.8.** *The Lie algebra of $Sp(1)$ is*

$$sp(1) = \mathit{span}\{(\mathbf{i}), (\mathbf{j}), (\mathbf{k})\},$$

*so* $\dim(Sp(1)) = 3$.

**Proof.** The path $\gamma_1(t) = (\cos(t) + \sin(t)\mathbf{i})$ in $Sp(1)$ satisfies $\gamma_1(0) = I$ and $\gamma_1'(0) = (\mathbf{i})$, so $\mathbf{i} \in sp(1)$. Similarly, $\gamma_2(t) = (\cos(t) + \sin(t)\mathbf{j})$ and $\gamma_3(t) = (\cos(t) + \sin(t)\mathbf{k})$ have initial velocities $\gamma_2'(0) = (\mathbf{j})$ and $\gamma_3'(0) = (\mathbf{k})$. So $\text{span}\{(\mathbf{i}), (\mathbf{j}), (\mathbf{k})\} \subset sp(1)$.

For the other inclusion, let $\gamma(t) = (a(t) + b(t)\mathbf{i} + c(t)\mathbf{j} + d(t)\mathbf{k})$ be a differentiable path in $Sp(1)$ with $\gamma(0) = I = (1)$. Since

$$|\gamma(t)|^2 = a(t)^2 + b(t)^2 + c(t)^2 + d(t)^2 = 1,$$

the value $a(0) = 1$ must be a local maximum of $a(t)$, so $a'(0) = 0$. Therefore $\gamma'(0) \in \text{span}\{(\mathbf{i}), (\mathbf{j}), (\mathbf{k})\}$. $\square$

In Figure 1, the circle group $U(1)$ and its Lie algebra are pictured on the left. The right image inaccurately represents $Sp(1)$ as $S^2 \subset \mathbb{R}^3$ rather than $S^3 \subset \mathbb{R}^4$, but is still a useful picture to keep in mind.

We end this section by describing the Lie algebras of the special linear groups.

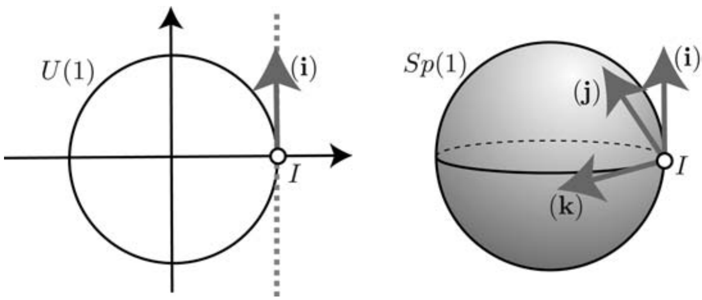



**Figure 1.** The Lie algebras of $U(1)$ and $Sp(1)$.

**Theorem 5.9.** *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. The Lie algebra, $sl_n(\mathbb{K})$, of $SL_n(\mathbb{K})$ is:*

$$sl_n(\mathbb{K}) = \{A \in M_n(\mathbb{K}) \mid trace(A) = 0\}.$$

*In particular,* $\dim(SL_n(\mathbb{R})) = n^2 - 1$ *and* $\dim(SL_n(\mathbb{C})) = 2(n^2 - 1)$.

The proof relies on the important fact that the trace is the derivative of the determinant; more precisely,

**Lemma 5.10.** *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. If $\gamma : (-\epsilon, \epsilon) \to M_n(\mathbb{K})$ is differentiable and $\gamma(0) = I$, then*

$$\frac{d}{dt}\Big|_{t=0} \det(\gamma(t)) = trace(\gamma'(0)).$$

**Proof.** Using the notation of Equation 1.5,

$$\begin{aligned}
&\frac{d}{dt}\Big|_{t=0} \det(\gamma(t)) \\
&\quad = \frac{d}{dt}\Big|_{t=0} \sum_{j=1}^{n} (-1)^{j+1} \cdot \gamma(t)_{1j} \cdot \det(\gamma(t)[1,j]) \\
&\quad = \sum_{j=1}^{n} (-1)^{j+1} \Big( \gamma'(0)_{1j} \cdot \det(\gamma(0)[1,j]) \\
&\qquad\qquad + \gamma(0)_{1j} \cdot \frac{d}{dt}\Big|_{t=0} \det(\gamma(t)[1,j]) \Big) \\
&\quad = \gamma'(0)_{11} + \frac{d}{dt}\Big|_{t=0} \det(\gamma(t)[1,1]) \qquad \text{(since } \gamma(0) = I\text{)}.
\end{aligned}$$

Re-applying the above argument to compute $\frac{d}{dt}|_{t=0}\det(\gamma(t)[1,1])$ and repeating $n$ times gives:

$$\frac{d}{dt}\bigg|_{t=0}\det(\gamma(t)) = \gamma'(0)_{11} + \gamma'(0)_{22} + \cdots + \gamma'(0)_{nn}.$$

□

**Proof of Theorem 5.9.** If $\gamma : (-\epsilon, \epsilon) \to SL_n(\mathbb{K})$ is differentiable with $\gamma(0) = I$, the lemma implies that $\text{trace}(\gamma'(0)) = 0$. This proves that every matrix in $sl_n(\mathbb{K})$ has trace zero.

On the other hand, suppose $A \in M_n(\mathbb{K})$ has trace zero. The path $\gamma(t) = I + tA$ satisfies $\gamma(0) = I$ and $\gamma'(0) = A$, but this path is not in $SL_n(\mathbb{K})$. Define $\alpha(t)$ as the result of multiplying each entry in the first row of $\gamma(t)$ by $1/\det(\gamma(t))$. Notice that $\alpha(t)$ is a differentiable path in $SL_n(\mathbb{K})$ with $\alpha(0) = I$. Further, since $\text{trace}(A) = 0$, it is straightforward to show that $\alpha'(0) = A$ (Exercise 5.2). This proves that every trace-zero matrix is in $sl_n(\mathbb{K})$. An alternative proof is to choose $\alpha(t)$ to be a one-parameter group, which will be introduced in the next chapter. □

## 3. Lie algebra vectors as vector fields

A **vector field** on $\mathbb{R}^m$ means a continuous function $F : \mathbb{R}^m \to \mathbb{R}^m$. By picturing $F(v)$ as a vector drawn at $v \in \mathbb{R}^m$, we think of a vector field as associating a vector to each point of $\mathbb{R}^m$.

If $A \in M_n(\mathbb{K})$, then $R_A : \mathbb{K}^n \to \mathbb{K}^n$ is a vector field on $\mathbb{K}^n$ ($= \mathbb{R}^n$, $\mathbb{R}^{2n}$ or $\mathbb{R}^{4n}$). The vector fields on $\mathbb{R}^2$ associated to the matrices $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ are shown in Figure 2.

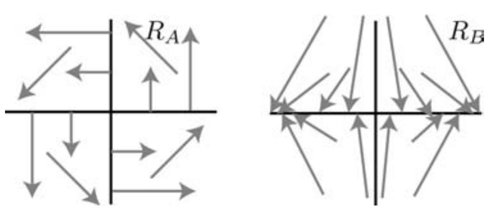



**Figure 2.** Vector fields on $\mathbb{R}^2$ associated to the matrices $A$ and $B$.

Elements of $GL_n(\mathbb{K})$ are thought of as linear transformations of $\mathbb{K}^n$ (by the correspondence $A \leftrightarrow R_A$); therefore, a differentiable path $\gamma : (-\epsilon, \epsilon) \to GL_n(\mathbb{K})$ should be regarded as a *one-parameter family* of linear transformations of $\mathbb{K}^n$. How does this family act on a single vector $X \in \mathbb{K}^n$? To decide this, let $\sigma(t) = R_{\gamma(t)}(X)$, which is a differentiable path in $\mathbb{K}^n$. If $\gamma(0) = I$, then $\sigma(0) = X$. By the product rule (which holds also for non-square matrices),

$$\sigma'(0) = R_{\gamma'(0)}(X).$$

We can therefore think of $R_{\gamma'(0)}$ as a vector field on $\mathbb{K}^n$ whose value at any $X \in \mathbb{K}^n$ tells the direction $X$ is initially moved by the family of linear transformations corresponding to the path $\gamma(t)$. In this way, it is often useful to visualize an element $\gamma'(0)$ of the Lie algebra of a matrix group $G \subset GL_n(\mathbb{K})$ as represented by the vector field $R_{\gamma'(0)}$ on $\mathbb{K}^n$.

For example, consider the path $\gamma(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$ in $SO(2)$. Its initial tangent vector, $A = \gamma'(0) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, lies in the Lie algebra, $so(2)$, of $SO(2)$. In fact, $so(2) = \text{span}\{A\}$. The vector field $R_A$ in Figure 2 illustrates how this family of rotations initially moves individual points of $\mathbb{R}^2$. The rotating action of the family $\gamma(t)$ of transformations is clearly manifested in the vector field $R_A$.

Next look at the graph of $R_B$ in Figure 2. Can you see from this graph why $B$ is not in the Lie algebra of $SO(2)$? If $\gamma(t)$ is a path in $GL_2(\mathbb{R})$ with $\gamma(0) = I$ and $\gamma'(0) = B$, then for small $t$, $R_{\gamma(t)} : \mathbb{R}^2 \to \mathbb{R}^2$ does not preserve norms. Which $X \in \mathbb{R}^2$ have initially increasing norms, and which are initially decreasing?

The vector field $R_A$ has an important visual property that $R_B$ lacks: the vector at any point is perpendicular to that point. By the above visual reasoning, we expect that for general $A \in M_n(\mathbb{R})$, if the vector field $R_A$ lacks this property, then $A$ could not lie in the Lie algebra, $so(n)$, of $SO(n)$. We could promote this visual reasoning to a careful proof without too much work (Exercise 5.11), but instead we use a cleaner, purely algebraic proof in the next section.

## 4. The Lie algebras of the orthogonal groups

The set $o_n(\mathbb{K}) = \{A \in M_n(\mathbb{K}) \mid A + A^* = 0\}$

... is denoted $so(n)$ and called the **skew-symmetric** matrices if $\mathbb{K} = \mathbb{R}$.

... is denoted $u(n)$ and called the **skew-hermitian** matrices if $\mathbb{K} = \mathbb{C}$.

... is denoted $sp(n)$ and called the **skew-symplectic** matrices if $\mathbb{K} = \mathbb{H}$.

We will prove that $o_n(\mathbb{K})$ is the Lie algebra of $\mathcal{O}_n(\mathbb{K})$. The condition $A = -A^*$ means that $A_{ij} = -\overline{A}_{ji}$ for all $i, j = 1...n$. So, the entries below the diagonal are determined by the entries above, and the diagonal entries are purely imaginary (which means zero if $\mathbb{K} = \mathbb{R}$). For example,

$$
\begin{aligned}
(5.2)\ u(2) \quad &= \quad \left\{ \begin{pmatrix} a\mathbf{i} & b + c\mathbf{i} \\ -b + c\mathbf{i} & d\mathbf{i} \end{pmatrix} \middle| a, b, c, d \in \mathbb{R} \right\} \\
&= \quad \operatorname{span} \left\{ \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \mathbf{i} \\ \mathbf{i} & 0 \end{pmatrix}, \begin{pmatrix} \mathbf{i} & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{i} \end{pmatrix} \right\},
\end{aligned}
$$

which is a 4-dimensional $\mathbb{R}$-subspace of $M_2(\mathbb{C})$, but not a $\mathbb{C}$-subspace. Similarly,

$$
sp(2) = \left\{ \begin{pmatrix} a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k} & x + y\mathbf{i} + z\mathbf{j} + w\mathbf{k} \\ -x + y\mathbf{i} + z\mathbf{j} + w\mathbf{k} & a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k} \end{pmatrix} \middle| \begin{matrix} a_i, b_i, c_i, \\ x, y, z, w \end{matrix} \in \mathbb{R} \right\},
$$

and

$$
so(3) = \left\{ \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix} \middle| a, b, c \in \mathbb{R} \right\}.
$$

If $A \in so(n)$, then the vector field $R_A$ on $\mathbb{R}^n$ has the property discussed in the previous section: the vector at any point is perpendicular to that point. This follows from (3) below:

**Lemma 5.11.** *For $A \in M_n(\mathbb{K})$, the following are equivalent:*

(1) $A \in o_n(\mathbb{K})$.

(2) $\langle R_A(X), Y\rangle = -\langle X, R_A(Y)\rangle$ *for all* $X, Y \in \mathbb{K}^n$.

(3) *(assuming* $\mathbb{K} = \mathbb{R}$*)* $\langle R_A(X), X\rangle = 0$ *for all* $X \in \mathbb{R}^n$.

**Proof.** To see that (1) $\implies$ (2), notice that for all $i, j = 1...n$,

$$\langle R_A(e_i), e_j \rangle = A_{ij} = -\overline{A}_{ji} = -\overline{\langle R_A(e_j), e_i \rangle} = -\langle e_i, R_A(e_j) \rangle.$$

This verifies (2) for $X, Y$ chosen from the standard orthonormal basis of $\mathbb{K}^n$. It is straightforward to extend linearly to arbitrary $X, Y \in \mathbb{K}^n$. The proof that (2) $\implies$ (1) is similar.

Now assume that $\mathbb{K} = \mathbb{R}$. In this case, (2) $\implies$ (3) by letting $X = Y$. To see that (3) $\implies$ (2), notice that:

$$\begin{aligned} 0 &= \langle R_A(X+Y), X+Y \rangle \\ &= \langle R_A(X), X \rangle + \langle R_A(Y), Y \rangle + \langle R_A(X), Y \rangle + \langle R_A(Y), X \rangle \\ &= 0 + 0 + \langle R_A(X), Y \rangle + \langle R_A(Y), X \rangle. \end{aligned}$$

□

**Theorem 5.12.** *The Lie algebra of $\mathcal{O}_n(\mathbb{K})$ equals $o_n(\mathbb{K})$.*

**Proof.** Suppose $\gamma : (-\epsilon, \epsilon) \to \mathcal{O}_n(\mathbb{K})$ is differentiable with $\gamma(0) = I$. Using the product rule to differentiate both sides of

$$\gamma(t) \cdot \gamma(t)^* = I$$

gives $\gamma'(0) + \gamma'(0)^* = 0$, so $\gamma'(0) \in o_n(\mathbb{K})$. This demonstrates that $\mathfrak{g}(\mathcal{O}_n(\mathbb{K})) \subset o_n(\mathbb{K})$.

Proving the other inclusion means explicitly constructing a path in $\mathcal{O}_n(\mathbb{K})$ in the direction of any $A \in o_n(\mathbb{K})$. It is simpler and sufficient to do so for all $A$ in a basis of $o_n(\mathbb{K})$.

The natural basis of $so(n) = o_n(\mathbb{R})$ is the set

$$\{E_{ij} - E_{ji} | 1 \leq i < j \leq n\},$$

where $E_{ij}$ denotes the matrix with $ij$-entry 1 and other entries zero. For example,

$$\begin{aligned} so(3) &= \text{span}\{E_{12} - E_{21}, E_{13} - E_{31}, E_{23} - E_{32}\} \\ &= \text{span}\left\{ \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \right\}. \end{aligned}$$

The path

$$\gamma_{ij}(t) = I + (\sin t)E_{ij} - (\sin t)E_{ji} + (-1 + \cos t)(E_{ii} + E_{jj})$$

lies in $SO(n)$, has $\gamma_{ij}(0) = I$ and has initial direction

$$\gamma'_{ij}(0) = E_{ij} - E_{ji}.$$

$R_{\gamma_{ij}(t)} : \mathbb{R}^n \to \mathbb{R}^n$ rotates the subspace $\text{span}\{e_i, e_j\}$ by an angle $t$ and does nothing to the other basis vectors. For example, the path

$$\gamma_{13}(t) = \begin{pmatrix} \cos t & 0 & \sin t \\ 0 & 1 & 0 \\ -\sin t & 0 & \cos t \end{pmatrix}$$

in $SO(3)$ satisfies $\gamma'_{13}(0) = E_{13} - E_{31}$. This proves the theorem for $\mathbb{K} = \mathbb{R}$. We leave it to the reader in Exercise 5.1 to describe a natural basis of $u(n)$ and $sp(n)$ and construct a path tangent to each element of those bases. ☐

**Corollary 5.13.**

(1) $\dim(SO(n)) = \frac{n(n-1)}{2}$.

(2) $\dim(U(n)) = n^2$.

(3) $\dim(Sp(n)) = 2n^2 + n$.

**Proof.** The $n^2$ entries of an $n$ by $n$ matrix include $d$ below the diagonal, $d$ above the diagonal, and $n$ on the diagonal. So $n^2 = d + d + n$, which means $d = \frac{n^2 - n}{2}$. Skew-symmetric matrices have zeros on the diagonal, arbitrary real numbers above, and entries below determined by those above, so $\dim(so(n)) = d$. Skew-hermitian matrices have purely imaginary numbers on the diagonal and arbitrary complex numbers above the diagonal, so $\dim(u(n)) = 2d + n = n^2$. Skew-symplectic matrices have elements of the form $a\mathbf{i} + b\mathbf{j} + d\mathbf{k}$ along the diagonal and arbitrary quaternionic numbers above the diagonal, so $\dim(sp(n)) = 4d + 3n = 2n^2 + n$. ☐

## 5. Exercises

**Ex. 5.1.** Complete the proof of Theorem 5.12.

**Ex. 5.2.** In the proof of Theorem 5.9, verify that $\alpha'(0) = A$.

**Ex. 5.3.** Prove that the product rule holds for non-square matrices.

**Ex. 5.4.** In Figure 2, how can you see visually that $A$ and $B$ both lie in $sl_2(\mathbb{R})$? Remember that a real $2 \times 2$ matrix with determinant 1 preserves the areas of parallelograms.

**Ex. 5.5.** Describe the Lie algebra of the affine group (see Exercise 3.11).

**Ex. 5.6.** Describe the Lie algebra of $\text{Isom}(\mathbb{R}^n)$.

**Ex. 5.7.** Describe the Lie algebra of $UT_n(\mathbb{K})$ (see Exercise 4.12).

**Ex. 5.8.** Prove the Lie algebra of $\rho_n(GL_n(\mathbb{C})) \subset GL_{2n}(\mathbb{R})$ is equal to $\rho_n(gl_n(\mathbb{C}))$.

**Ex. 5.9.** Prove the Lie algebra of $\Psi_n(GL_n(\mathbb{H})) \subset GL_{2n}(\mathbb{C})$ is $\Psi_n(gl_n(\mathbb{H}))$.

**Ex. 5.10.** Prove that the tangent space to a matrix group $G$ at $A \in G$ is:

$$T_A(G) = \{BA \mid B \in \mathfrak{g}(G)\} = \{AB \mid B \in \mathfrak{g}(G)\}.$$

**Ex. 5.11.** Give a geometric proof of the fact at the end of Section 3.

**Ex. 5.12.** Give an example of a 2-dimensional matrix group.

**Ex. 5.13.** Is Lemma 5.10 true for $\mathbb{K} = \mathbb{H}$?

**Ex. 5.14.** Describe the Lie algebra of $SL_n(\mathbb{H})$.

**Ex. 5.15.** Is part 3 of Lemma 5.11 valid when $\mathbb{K} \in \{\mathbb{C}, \mathbb{H}\}$?

**Ex. 5.16.** Let $\gamma : (-\epsilon, \epsilon) \to GL_n(\mathbb{K})$ be a differentiable path. Prove that the inverse path $t \mapsto \gamma(t)^{-1}$ is differentiable.
*Hint: For* $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$*, use Cramer's rule.*

Chapter 6

# Matrix exponentiation

To prove Theorem 5.12, which said $\mathfrak{g}(\mathcal{O}_n(\mathbb{K})) = o_n(\mathbb{K})$, we constructed a differentiable path through the identity in $\mathcal{O}_n(\mathbb{K})$ in the direction of any $A$ in a basis of $o_n(\mathbb{K})$. Our paths were defined with sines and cosines and seemed natural because they corresponded to families of rotations in certain planes. On the other hand, the paths we constructed to prove Theorem 5.9 (verifying the Lie algebra of $SL_n(\mathbb{K})$) seemed less natural. In general, is there a "best" path in the direction of any $A \in gl_n(\mathbb{K})$, and is this best path guaranteed to be contained in any matrix group $G \subset GL_n(\mathbb{K})$ to which $A$ is a tangent vector? In this chapter, we construct optimal paths, which are called *one-parameter groups* and are defined in terms of *matrix exponentiation*. We begin the chapter with preliminary facts about series, which are necessary to understand matrix exponentiation.

## 1. Series in $\mathbb{K}$

We say that a series

$$\sum a_l = a_0 + a_1 + a_2 + \cdots$$

of elements $a_l \in \mathbb{K}$ **converges** if the corresponding sequence of partial sums

$$\{a_0, a_0 + a_1, a_0 + a_1 + a_2, ...\}$$

converges to some $a \in \mathbb{K}$. Here we are regarding $\mathbb{K}$ as $\mathbb{R}$, $\mathbb{R}^2$, or $\mathbb{R}^4$, and "convergence" means in the sense of Definition 4.6. In this case, we write $\sum a_l = a$. The series $\sum a_l$ is said to **converge absolutely** if $\sum |a_l|$ converges.

**Proposition 6.1.** *If $\sum a_l$ converges absolutely, then it converges.*

**Proof.** By the triangle inequality,

$$\left| \sum_{l=l_1}^{l_2} a_l \right| \leq \sum_{l=l_1}^{l_2} |a_l|.$$

The right side of this inequality is the distance between the $l_2$-th and the $l_1$-th partial sums of $\sum |a_l|$. The left side equals the distance between the $l_2$-th and the $l_1$-th partial sums of $\sum a_l$. If $\sum |a_l|$ converges, then its sequence of partial sums is Cauchy, so the inequality implies that the sequence of partial sums of $\sum a_l$ is also Cauchy and therefore convergent by Proposition 4.9. $\square$

One expects that the product of two series can be calculated by "infinitely distributing" and organizing terms by the sum of the indices, as in:

$$\begin{aligned}(a_0 + a_1 + a_2 + \cdots)&(b_0 + b_1 + b_2 + \cdots) \\ &= (a_0b_0) + (a_0b_1 + a_1b_0) + (a_0b_2 + a_1b_1 + a_2b_0) + \cdots.\end{aligned}$$

This manipulation is justified by the following fact, which is proven in most analysis textbooks:

**Proposition 6.2.** *Suppose that $\sum a_l$ and $\sum b_l$ both converge, at least one absolutely. Let $c_l = \sum_{k=0}^{l} a_k b_{l-k}$. Then $\sum c_l = (\sum a_l)(\sum b_l)$.*

A **power series** means a "formal infinite-degree polynomial," that is, an expression of the form:

$$f(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + \cdots$$

with coefficients $c_i \in \mathbb{K}$. When the variable $x$ is assigned a value in $\mathbb{K}$, the result is a series which may or may not converge. The **domain** of $f$ means the set of all $x \in \mathbb{K}$ for which the series $f(x)$ converges. The next proposition says that the domain of any power series is a ball about the origin (possibly including some of its boundary and possibly with a radius of zero or infinity).

**Proposition 6.3.** *For any power series there exists an $R \in [0, \infty]$ (called its **radius of convergence**) such that $f(x)$ converges absolutely if $|x| < R$ and diverges if $|x| > R$.*

When $R = 0$, the series converges only at $x = 0$. When $R = \infty$, the series converges for all $x \in \mathbb{K}$.

**Proof.** The **root test** says that a series $\sum a_n$ converges absolutely if

$$\alpha = \limsup_{n\to\infty}(|a_n|)^{1/n}$$

is less than one, and diverges if $\alpha$ is greater than one. Even when $\mathbb{K} \in \{\mathbb{C}, \mathbb{H}\}$, this is essentially a statement about series of positive real numbers, so the $\mathbb{K} = \mathbb{R}$ proof found in any calculus textbook needs no alteration. In the series obtained by substituting $x \in \mathbb{K}$ into the power series $f(x) = \sum c_n x^n$,

$$\alpha = |x| \limsup_{n\to\infty}(|c_n|)^{1/n},$$

so the proposition holds with

$$R = \left(\limsup_{n\to\infty}(|c_n|)^{1/n}\right)^{-1}.$$

The interpretations of the extreme cases are: if $\limsup_{n\to\infty}(|c_n|)^{1/n}$ equals zero, then $R = \infty$, and if it equals $\infty$, then $R$ equals zero. □

In future applications, we will often restrict a power series to the real numbers in its domain. Such a restriction can be differentiated term-by-term as follows:

**Proposition 6.4.** *Let $f(x) = c_0 + c_1 x + c_2 x^2 + \cdots$ be a power series with radius of convergence $R$. The restriction of $f$ to the real numbers in its domain, $f : (-R, R) \to \mathbb{K}$, is a differentiable path in $\mathbb{K}$ with derivative $f'(x) = c_1 + 2c_2 x + 3c_3 x^2 + \cdots$.*

**Proof.** The case $\mathbb{K} = \mathbb{R}$ is familiar from calculus, and the general case follows immediately from the real case. □

## 2. Series in $M_n(\mathbb{K})$

We will also study series of matrices. We say that a series

$$\sum A_l = A_0 + A_1 + A_2 + \cdots$$

of elements $A_l \in M_n(\mathbb{K})$ converges (absolutely) if for each $i, j$ the series $(A_0)_{ij} + (A_1)_{ij} + (A_2)_{ij} + \cdots$ converges (absolutely) to some $A_{ij} \in \mathbb{K}$. In this case, we write $\sum A_l = A$.

Proposition 6.2 generalizes to series of matrices:

**Proposition 6.5.** *Suppose that $\sum A_l$ and $\sum B_l$ both converge, at least one absolutely. Let $C_l = \sum_{k=0}^{l} A_k B_{l-k}$. Then,*

$$\sum C_l = (\sum A_l)(\sum B_l).$$

The proof of Proposition 6.5 is left for Exercise 6.1. The idea is to use Proposition 6.2 to prove that for all $i, j$,

$$\left(\sum C_l\right)_{ij} = \left(\left(\sum A_l\right)\left(\sum B_l\right)\right)_{ij}.$$

A power series $f(x) = c_0 + c_1 x + c_2 x^2 + \cdots$ with coefficients $c_i \in \mathbb{K}$ can be evaluated on a matrix $A \in M_n(\mathbb{K})$. The result is a series in $M_n(\mathbb{K})$:

$$f(A) = c_0 I + c_1 A + c_2 A^2 + \cdots .$$

**Proposition 6.6.** *Let $f(x) = c_0 + c_1 x + c_2 x^2 + \cdots$ be a power series with coefficients $c_i \in \mathbb{K}$ with radius of convergence $R$. If $A \in M_n(\mathbb{K})$ satisfies $|A| < R$, then $f(A) = c_0 I + c_1 A + c_2 A^2 + \cdots$ converges absolutely.*

Remember that $|A|$ denotes the Euclidean norm on $M_n(\mathbb{K})$ regarded as $\mathbb{R}^{n^2}$, $\mathbb{R}^{2n^2}$ or $\mathbb{R}^{4n^2}$. For example,

$$\left| \begin{pmatrix} a + b\mathbf{i} & c + d\mathbf{i} \\ e + f\mathbf{i} & g + h\mathbf{i} \end{pmatrix} \right| = \sqrt{a^2 + b^2 + c^2 + d^2 + e^2 + f^2 + g^2 + h^2}.$$

The proof of Proposition 6.6 will require an important lemma:

**Lemma 6.7.** *For all $X, Y \in M_n(\mathbb{K})$, $|XY| \leq |X| \cdot |Y|$.*

**Proof.** The proof depends on Proposition 3.7, the Schwarz inequality. For all indices $i, j$,

$$\begin{aligned}
|(XY)_{ij}|^2 &= \left|\sum_{l=1}^n X_{il}Y_{lj}\right|^2 = \left|\langle(\text{row } i \text{ of } X), (\text{column } j \text{ of } \overline{Y})^T\rangle\right|^2 \\
&\leq \left|(\text{row } i \text{ of } X)\right|^2 \cdot \left|(\text{column } j \text{ of } \overline{Y})^T\right|^2 \\
&= \left(\sum_{l=1}^n |X_{il}|^2\right) \cdot \left(\sum_{l=1}^n |Y_{lj}|^2\right).
\end{aligned}$$

Summing over all indices $i, j$ gives:

$$\begin{aligned}
|XY|^2 &= \sum_{i,j=1}^n |(XY)_{ij}|^2 \leq \sum_{i,j=1}^n \left(\left(\sum_{l=1}^n |X_{il}|^2\right) \cdot \left(\sum_{l=1}^n |Y_{lj}|^2\right)\right) \\
&= \left(\sum_{i,j=1}^n |X_{ij}|^2\right) \cdot \left(\sum_{i,j=1}^n |Y_{ij}|^2\right) = |X|^2|Y|^2.
\end{aligned}$$

□

**Proof of Proposition 6.6.** For any indices $i, j$, we must prove that

$$|(c_oI)_{ij}| + |(c_1A)_{ij}| + |(c_2A^2)_{ij}| + \cdots$$

converges. The $l$th term of this series satisfies:

$$|(c_lA^l)_{ij}| \leq |c_lA^l| = |c_l| \cdot |A^l| \leq |c_l| \cdot |A|^l.$$

Since $|A|$ is less than the radius of convergence of $f$, the result follows. □

When the power series of the function $f(x) = e^x$ is applied to a matrix $A \in M_n(\mathbb{K})$, the result is called **matrix exponentiation**:

$$e^A = I + A + (1/2!)A^2 + (1/3!)A^3 + (1/4!)A^4 + \cdots .$$

The radius of convergence of this power series is $\infty$, so by Proposition 6.6, $e^A$ converges absolutely for all $A \in M_n(\mathbb{K})$. As you might guess from its appearance as the chapter title, matrix exponentiation is a central idea in the study of matrix groups.

## 3. The best path in a matrix group

In this section, we use matrix exponentiation to construct canonical "best paths" in each given direction of a matrix group. Let's begin with a simple example. Figure 1 illustrates the vector field associated to

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in so(2).$$

What is the most natural differentiable path $\gamma(t)$ in $SO(2)$ with $\gamma(0) = I$ and $\gamma'(0) = A$? The choice $\gamma(t) = I + tA$ seems natural, but is not in $SO(2)$. Every path in $SO(2)$ with the desired initial conditions has the form:

$$\gamma(t) = \begin{pmatrix} \cos f(t) & \sin f(t) \\ -\sin f(t) & \cos f(t) \end{pmatrix},$$

where $f(t)$ is a differentiable function with $f(0) = 0$ and $f'(0) = 1$. The choice $f(t) = t$ is clearly the most natural choice; in fact, this in the unique choice that gives $\gamma$ the following natural property: for every $X \in \mathbb{R}^2$, the path $\alpha(t) = R_{\gamma(t)}(X)$ is an *integral curve* of the vector field $R_A$. This means that the vector field $R_A$ tells the direction that $X$ is moved by the family of linear transformations associated to $\gamma(t)$ for all time (rather than just initially at $t = 0$). More precisely,

**Definition 6.8.** *A path $\alpha : (-\epsilon, \epsilon) \to \mathbb{R}^m$ is called an **integral curve** of a vector field $F : \mathbb{R}^m \to \mathbb{R}^m$ if $\alpha'(t) = F(\alpha(t))$ for all $t \in (-\epsilon, \epsilon)$.*

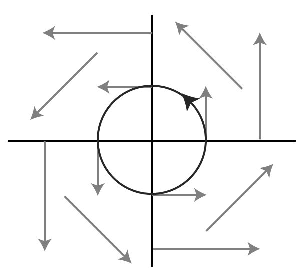


**Figure 1.** An integral curve of $R_A$.

For the matrix $A$ above, the integral curves of $R_A$ are (segments of) circles centered at the origin, parameterized counterclockwise with speed one radian per unit time.

More generally, if $A \in gl_n(\mathbb{K})$, we would like to find the "most natural" path $\gamma(t)$ in $GL_n(\mathbb{K})$ with $\gamma(0) = I$ and $\gamma'(0) = A$. We will attempt to choose $\gamma(t)$ such that for all $X \in \mathbb{K}^n$, the path

$$t \mapsto R_{\gamma(t)}(X)$$

is an integral curve of $R_A$. You might find it surprising that a single path $\gamma(t)$ will work for all choices of $X$.

The trick is to find a power series expression for the integral curve $\alpha(t)$ of $R_A$ beginning at $\alpha(0) = X$. We contrive coefficients $c_i \in \mathbb{K}^n$ such that the path $\alpha : \mathbb{R} \to \mathbb{K}^n$ defined by the power series

$$\alpha(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3 + \cdots$$

is an integral curve of $R_A$ with $\alpha(0) = c_0 = X$. Being an integral curve means that $\alpha'(t)$ (which is $c_1 + 2c_2 t + 3c_3 t^2 + \cdots$) equals $R_A(\alpha(t))$ (which is $c_0 A + c_1 tA + c_2 t^2 A + c_3 t^3 A + \cdots$). So we want:

$$(c_1 + 2c_2 t + 3c_3 t^2 + \cdots) = (c_0 A + c_1 tA + c_2 t^2 A + c_3 t^3 A + \cdots).$$

Equating coefficients of corresponding powers of $t$ gives the recursive formula $lc_l = c_{l-1}A$. Together with the initial condition $c_0 = X$, this gives the explicit formula $c_l = \frac{1}{l!} XA^l$, so the integral curve is:

$$\begin{aligned} \alpha(t) &= X + XtA + \frac{X}{2!}(tA)^2 + \frac{X}{3!}(tA)^3 + \cdots \\ &= Xe^{tA} = R_{e^{tA}}(X). \end{aligned}$$

In summary, the path $\gamma(t) = e^{tA}$ has the desired property that for all $X \in \mathbb{K}^n$, $\alpha(t) = R_{\gamma(t)}(X)$ is an integral curve of $R_A$. Instead of deriving this result, it would have been faster to verify it from scratch using:

**Proposition 6.9.** *Let $A \in gl_n(\mathbb{K})$. The path $\gamma : \mathbb{R} \to M_n(\mathbb{K})$ defined as $\gamma(t) = e^{tA}$ is differentiable, and $\gamma'(t) = A \cdot \gamma(t) = \gamma(t) \cdot A$.*

**Proof.** Each of the $n^2$ entries of

$$\gamma(t) = e^{tA} = I + tA + (1/2)t^2 A^2 + (1/6)t^3 A^3 + \cdots$$

is a power series in $t$, which by Proposition 6.4 can be termwise differentiated, giving:

$$\gamma'(t) = 0 + A + tA^2 + (1/2)t^2 A^3 + \cdots .$$

This equals $\gamma(t) \cdot A$ or $A \cdot \gamma(t)$ depending on whether you factor an $A$ out on the left or right. $\square$

There are two interesting interpretations of Proposition 6.9, the first of which we've already discussed:

**Proposition 6.10.** *Let $A \in M_n(\mathbb{K})$ and let $\gamma(t) = e^{tA}$.*

(1) *For all $X \in \mathbb{K}^n$, $\alpha(t) = R_{\gamma(t)}(X)$ is an integral curve of $R_A$. Also, $\alpha(t) = L_{\gamma(t)}(X)$ is an integral curve of $L_A$.*

(2) *$\gamma(t)$ is itself an integral curve of the vector field on $M_n(\mathbb{K})$ whose value at $g$ is $A \cdot g$ (and is also an integral curve of the vector field whose value at $g$ is $g \cdot A$).*

Both (1) and (2) follow immediately from Proposition 6.9. The two parts have different pictures and different uses. It is interesting that the left and right versions of part (2) can simultaneously be true, since the two vector fields on $M_n(\mathbb{K})$ do not agree. Evidently, they agree along the image of $\gamma$.

## 4. Properties of the exponential map

The exponential map $\exp : M_n(\mathbb{K}) \to M_n(\mathbb{K})$, which sends $A \mapsto e^A$, is a powerful tool for studying matrix groups. We have already seen that exp restricted to a real line is a "best path". In this section, we derive important algebraic properties of the exponential map, which further justify our use of the term "exponential".

**Proposition 6.11.** *If $AB = BA$, then $e^{A+B} = e^A \cdot e^B$.*

**Proof.** By Proposition 6.5,

$$\begin{aligned} e^A e^B &= (I + A + (1/2)A^2 + \cdots)(I + B + (1/2)B^2 + \cdots) \\ &= I + (A+B) + ((1/2)A^2 + AB + (1/2)B^2) + \cdots . \end{aligned}$$

On the other hand,

$$\begin{aligned} e^{A+B} &= I + (A+B) + (1/2)(A+B)^2 + \cdots \\ &= I + (A+B) + (1/2)(A^2 + AB + BA + B^2) + \cdots . \end{aligned}$$

Since $AB = BA$, the first terms of $e^A e^B$ equal the first terms of $e^{A+B}$. To verify that the pattern continues:

$$\begin{aligned} e^A e^B &= \left(\sum_{l=0}^{\infty} \frac{A^l}{l!}\right)\left(\sum_{l=0}^{\infty} \frac{B^l}{l!}\right) = \sum_{l=0}^{\infty}\sum_{k=0}^{l} \frac{A^k B^{l-k}}{k!(l-k)!} \\ &= \sum_{l=0}^{\infty} \frac{1}{l!} \sum_{k=0}^{l} \binom{l}{k} A^k B^{l-k} = \sum_{l=0}^{\infty} \frac{(A+B)^l}{l!}. \end{aligned}$$

The last equality uses the fact that $A$ and $B$ commute. □

Since most pairs of matrices do not commute, you might not expect Proposition 6.11 to have much use, except in the $n = 1$ case. Surprisingly, the proposition has many strong consequences, including every proposition in the remainder of this section.

**Proposition 6.12.** *For any $A \in M_n(\mathbb{K})$, $e^A \in GL_n(\mathbb{K})$. Therefore, matrix exponentiation is a map* $\exp : gl_n(\mathbb{K}) \to GL_n(\mathbb{K})$.

**Proof.** Since $A$ and $-A$ commute, $e^A \cdot e^{-A} = e^{A-A} = e^0 = I$, so $e^A$ has inverse $e^{-A}$. □

We will see later that the image of exp contains a neighborhood of $I$ in $M_n(\mathbb{K})$, so it may seem counterintuitive that this image misses all of the singular matrices.

**Proposition 6.13.** *If $A \in o_n(\mathbb{K})$, then $e^A \in \mathcal{O}_n(\mathbb{K})$.*

**Proof.** Since $A \in o_n(\mathbb{K})$, $A^* = -A$. Therefore,

$$e^A \left(e^A\right)^* = e^A e^{A^*} = e^A e^{-A} = e^{A-A} = e^0 = I.$$

So $e^A \in \mathcal{O}_n(\mathbb{K})$ by part (4) of Proposition 3.9. Exercise 6.2 asks you to verify that $(e^A)^* = e^{A^*}$, which was used in the first equality above. □

This proposition allows a cleaner proof of Theorem 5.12, which says that $o_n(\mathbb{K})$ is the Lie algebra of $\mathcal{O}_n(\mathbb{K})$. How? If $A \in o_n(\mathbb{K})$, then $\gamma(t) = e^{tA}$ is a differentiable path in $\mathcal{O}_n(\mathbb{K})$ with $\gamma'(0) = A$. This proves that $o_n(\mathbb{K}) \subset \mathfrak{g}(\mathcal{O}_n(\mathbb{K}))$, which was the more difficult inclusion to verify.

Since $SU(n) = U(n) \cap SL_n(\mathbb{C})$, one expects the Lie algebra $su(n)$ of $SU(n)$ to equal the set of trace-zero skew-hermitian matrices:

**Proposition 6.14.** $su(n) = u(n) \cap sl_n(\mathbb{C})$.

The inclusion $su(n) \subset u(n) \cap sl_n(\mathbb{C})$ is trivial. For the other inclusion, we must construct a path in $SU(n)$ tangent to any $A \in u(n) \cap sl_n(\mathbb{C})$. The path $\gamma(t) = e^{tA}$ is contained in $U(n)$, but we have yet to verify that it is contained in $SL_n(\mathbb{C})$, which follows from:

**Lemma 6.15.** *Let* $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. *For any* $A \in M_n(\mathbb{K})$,

$$\det(e^A) = e^{trace(A)}.$$

**Proof.** Let $f(t) = \det(e^{tA})$. Its derivative is:

$$\begin{aligned}
f'(t) &= \lim_{h\to 0}(1/h)(\det(e^{(t+h)A}) - \det(e^{tA})) \\
&= \lim_{h\to 0}(1/h)(\det(e^{tA}e^{hA}) - \det(e^{tA})) \\
&= \lim_{h\to 0}(1/h)(\det(e^{tA}))(\det(e^{hA}) - 1) \\
&= (\det(e^{tA}))\lim_{h\to 0}(1/h)(\det(e^{hA}) - 1) \\
&= f(t)\left.\frac{d}{dt}\right|_{t=0}\det\left(e^{tA}\right) \\
&= f(t)\cdot \mathrm{trace}(A).
\end{aligned}$$

The last equality follows from Lemma 5.10. Since $f(0) = 1$ and

$$f'(t) = f(t)\cdot \mathrm{trace}(A),$$

the unique solution for $f$ is $f(t) = e^{t\cdot \mathrm{trace}(A)}$. In particular,

$$f(1) = \det(e^A) = e^{\mathrm{trace}(A)}.$$

□

For $A \in gl_n(\mathbb{K})$, we have verified that the path $\gamma(t) = e^{tA}$ has several geometric and analytic properties. It also has an important algebraic property; namely, its image is a subgroup of $GL_n(\mathbb{K})$. To elaborate on this comment, we make the following definition, in which $(\mathbb{R}, +)$ denotes the group of real numbers under the operation of addition.

**Definition 6.16.** *A **one-parameter group** in a matrix group $G$ is a differentiable group-homomorphism $\gamma : (\mathbb{R}, +) \to G$.*

"Homomorphism" means that $\gamma(t_1 + t_2) = \gamma(t_1)\gamma(t_2)$. A one-parameter group is both an algebraic object (a homomorphism) and a geometric object (a differentiable path). The interplay between algebra and geometry is what makes matrix groups so rich in structure.

**Proposition 6.17.**

(1) *For every $A \in gl_n(\mathbb{K})$, $\gamma(t) = e^{tA}$ is a one-parameter group.*

(2) *Every one-parameter group in $GL_n(\mathbb{K})$ has the description $\gamma(t) = e^{tA}$ for some $A \in gl_n(\mathbb{K})$.*

**Proof.** Part (1) follows from Proposition 6.11, since:

$$\gamma(t_1 + t_2) = e^{t_1 A + t_2 A} = e^{t_1 A} e^{t_2 A} = \gamma(t_1)\gamma(t_2).$$

Notice in particular that $\gamma(t) \cdot \gamma(-t) = I$, which shows that

$$\gamma(t)^{-1} = \gamma(-t).$$

For part (2), suppose $\gamma(t)$ is a one-parameter group in $GL_n(\mathbb{K})$. Let $A = \gamma'(0)$. Notice that for all $t \in \mathbb{R}$,

$$\gamma'(t) = \lim_{h \to 0} \frac{1}{h}(\gamma(t+h) - \gamma(t)) = \gamma(t) \lim_{h \to 0} \frac{1}{h}(\gamma(h) - I) = \gamma(t)A.$$

Since $\gamma'(t) = \gamma(t)A$, we suspect (by comparing to Proposition 6.9) that $\gamma(t) = e^{tA}$. This is verified by applying the product rule:

$$\begin{aligned} \frac{d}{dt}(\gamma(t)e^{-tA}) &= \gamma'(t)e^{-tA} + \gamma(t)\frac{d}{dt}(e^{-tA}) \\ &= \gamma(t)Ae^{-tA} - \gamma(t)Ae^{-tA} = 0. \end{aligned}$$

So $\gamma(t)e^{-tA} = I$, which implies that $\gamma(t) = e^{tA}$. $\square$

Finally, we describe how conjugation and exponentiation relate:

**Proposition 6.18.** *For all $A, B \in M_n(\mathbb{K})$ with $A$ invertible,*

$$e^{ABA^{-1}} = Ae^{B}A^{-1}.$$

**Proof.**

$$
\begin{aligned}
Ae^B A^{-1} &= A(I + B + (1/2)B^2 + (1/6)B^3 + \cdots)A^{-1} \\
&= I + ABA^{-1} + (1/2)AB^2A^{-1} + (1/6)AB^3A^{-1} + \cdots \\
&= I + ABA^{-1} + (1/2)(ABA^{-1})^2 + (1/6)(ABA^{-1})^3 + \cdots \\
&= e^{ABA^{-1}}.
\end{aligned}
$$

□

## 5. Exercises

**Ex. 6.1.** Prove Proposition 6.5.

**Ex. 6.2.** Prove that $(e^A)^* = e^{A^*}$ for all $A \in M_n(\mathbb{K})$.

**Ex. 6.3.**

(1) Let $A = \text{diag}(a_1, a_2, ..., a_n) \in M_n(\mathbb{R})$. Calculate $e^A$. Using this, give a simple proof that $\det(e^A) = e^{\text{trace}(A)}$ when $A$ is diagonal.

(2) Give a simple proof that $\det(e^A) = e^{\text{trace}(A)}$ when $A$ is conjugate to a diagonal matrix.

**Ex. 6.4.** Let $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Calculate $e^{tA}$ for arbitrary $t \in \mathbb{R}$.

**Ex. 6.5.** Can a one-parameter group ever cross itself?

**Ex. 6.6.** Describe all one-parameter groups in $GL_1(\mathbb{C})$. Draw several in the $xy$-plane.

**Ex. 6.7.** Let $G = \left\{ \begin{pmatrix} x & y \\ 0 & 1 \end{pmatrix} \in GL_2(\mathbb{R}) \middle| x > 0 \right\}$. Describe the one-parameter groups in $G$, and draw several in the $xy$-plane.

**Ex. 6.8.** Visually describe the path $\gamma(t) = e^{t\mathbf{j}}$ in $Sp(1) \cong S^3$.

**Ex. 6.9.** Let $A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \in gl_2(\mathbb{R})$. Calculate $e^A$. (*Hint: Write $A$ as the sum of two commuting matrices.*) Draw the vector field $R_A$ when $a = 1$ and $b = 2$, and sketch some integral curves.

**Ex. 6.10.** Repeat the previous problem with $A = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$.

**Ex. 6.11.** When $A$ is in the Lie algebra of $UT_n(\mathbb{K})$, prove that $e^A \in UT_n(\mathbb{K})$ (see Exercise 4.12).

**Ex. 6.12.** When $A$ is in the Lie algebra of $\text{Isom}(\mathbb{R}^n)$, prove that $e^A \in \text{Isom}(\mathbb{R}^n)$.

**Ex. 6.13.** Describe the one-parameter groups of $\text{Trans}(\mathbb{R}^n)$.

**Ex. 6.14.** The multiplicative group of positive real numbers can be identified with the subgroup: $G = \{(a) \in GL_1(\mathbb{R}) \mid a > 0\}$. Given $A$ in the Lie algebra of $G$, describe the vector field on $G$ associated to $A$, as part (2) of Proposition 6.10. Solve for the integral curve of this vector field.

Chapter 7

# Matrix groups are manifolds

In this chapter we prove two crucial facts about how the exponential map restricts to a Lie algebra. For $r > 0$, denote

$$B_r = \{W \in M_n(\mathbb{K}) \mid |W| < r\}.$$

**Theorem 7.1.** *Let $G \subset GL_n(\mathbb{K})$ be a matrix group, with Lie algebra $\mathfrak{g} \subset gl_n(\mathbb{K})$.*

(1) *For all $X \in \mathfrak{g}$, $e^X \in G$.*

(2) *For sufficiently small $r > 0$, $V = \exp(B_r \cap \mathfrak{g})$ is a neighborhood of $I$ in $G$, and the restriction $\exp : B_r \cap \mathfrak{g} \to V$ is a homeomorphism.*

Part (1) says that if a one-parameter group in $GL_n(\mathbb{K})$ begins tangent to a matrix group $G$, then it lies entirely in $G$. In the previous chapter, we verified (1) when $G \in \{GL_n(\mathbb{K}), \mathcal{O}_n(\mathbb{K}), SL_n(\mathbb{R}), SL_n(\mathbb{C}), SU(n)\}$. However, the proofs were different in each case, and new ideas are needed in this chapter to generalize to arbitrary matrix groups.

Part (2) has not yet been verified for any familiar matrix groups. We will actually prove the stronger statement that $\exp : B_r \cap \mathfrak{g} \to V$ is a *diffeomorphism* (which will be defined in this chapter, but roughly means a differentiable homeomorphism).

A beautiful corollary of Theorem 7.1 is that every matrix group is a manifold, which we will carefully define in this chapter. Roughly, a manifold is a nice subset of Euclidean space; at every point $p$ its tangent space is a subspace, and a neighborhood of $p$ is diffeomorphic to a neighborhood of 0 in the tangent space. Manifolds are central to modern mathematics. Their investigation is the starting point of several branches of geometry.

## 1. Analysis background

In this section, we review some concepts from analysis that are necessary to prove Theorem 7.1.

Let $U \subset \mathbb{R}^m$ be an open set. Any function $f : U \to \mathbb{R}^n$ can be thought of as $n$ separate functions; we write $f = (f_1, ..., f_n)$, where each $f_i : U \to \mathbb{R}$. For example, the function $f : \mathbb{R}^2 \to \mathbb{R}^3$ defined as

$$f(x, y) = (\sin(x + y), e^{xy}, x^2 - y^3)$$

splits as $f_1(x, y) = \sin(x + y)$, $f_2(x, y) = e^{xy}$ and $f_3(x, y) = x^2 - y^3$.

Let $p \in U$ and let $v \in \mathbb{R}^m$. The **directional derivative** of $f$ in the direction $v$ at $p$ is defined as:

$$df_p(v) = \lim_{t \to 0} \frac{f(p + tv) - f(p)}{t},$$

if this limit exists.

The directional derivative can be interpreted visually by considering the straight line $\gamma(t) = p+tv$ in $\mathbb{R}^m$. If the initial velocity vector of the image path $(f \circ \gamma)(t) = f(p+tv)$ in $\mathbb{R}^n$ exists, it is called $df_p(v)$. In other words, $df_p(v)$ approximates where $f$ sends points near $p$ in the direction of $v$; see Figure 1.

The directional derivatives of the component functions $\{f_1, ..., f_n\}$ in the directions of the standard orthonormal basis vectors $\{e_1, ..., e_m\}$ of $\mathbb{R}^m$ are called **partial derivatives** of $f$ and are denoted as:

$$\frac{\partial f_i}{\partial x_j}(p) = d(f_i)_p(e_j).$$

They measure the rates at which the component functions change in the coordinate directions. For fixed $\{i, j\}$, if $\frac{\partial f_i}{\partial x_j}(p)$ exists at each

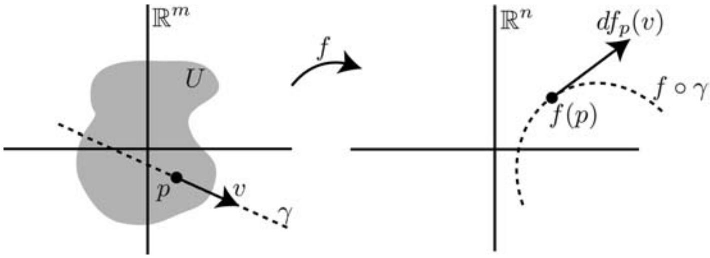



**Figure 1.** $df_p(v)$ is the initial velocity vector of the image under $f$ of the straight line $\gamma(t)$ in the direction of $v$.

$p \in U$, then $p \mapsto \frac{\partial f_i}{\partial x_j}(p)$ is another function from $U$ to $\mathbb{R}$; its partial derivatives (if they exist) are called second order partial derivatives of $f$, and so on.

The function $f$ is called "$C^r$ on $U$" if all $r^{\text{th}}$ order partial derivatives exist and are continuous on $U$, and $f$ is called **smooth** on $U$ if $f$ is $C^r$ on $U$ for all positive integers $r$. The following is proven in any real analysis textbook:

**Proposition 7.2.** *If $f$ is $C^1$ on $U$, then for all $p \in U$,*

(1) *$v \mapsto df_p(v)$ is a linear function from $\mathbb{R}^m$ to $\mathbb{R}^n$.*

(2) *$f(q) \approx f(p) + df_p(q-p)$ is a good approximation of $f$ near $p$ in the following sense: for any infinite sequence $\{q_1, q_2, ...\}$ of points in $\mathbb{R}^m$ converging to $p$,*

$$\lim_{t\to\infty} \frac{f(q_t) - f(p) - df_p(q_t - p)}{|q_t - p|} = 0.$$

Proposition 7.2 says that if $f$ is $C^1$ on $U$, then the directional derivatives of $f$ are well-behaved at any $p \in U$. It is useful to turn this conclusion into a definition.

**Definition 7.3.** *$f : \mathbb{R}^m \to \mathbb{R}^n$ is called **differentiable** at $p \in \mathbb{R}^m$ if $df_p(v)$ exists for every $v \in \mathbb{R}^m$, and properties (1) and (2) of Proposition 7.2 hold. In this case, the linear function $df_p$ is called the **derivative** of $f$ at $p$.*

Notice that it is not enough for all directional derivatives to exist at $p$; we require the function $v \mapsto df_p(v)$ to be linear and to approximate $f$ well near $p$ before we are willing to call $f$ differentiable at $p$ or to refer to the function $df_p$ as its derivative.

If $f$ is $C^1$ on a neighborhood of $p$, then $f$ is differentiable at $p$, and $df_p = L_A = R_{A^T}$, where $A \in M_{n,m}(\mathbb{R})$ is the matrix of all first order partial derivatives of $f$:

$$A = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(p) & \cdots & \frac{\partial f_1}{\partial x_m}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(p) & \cdots & \frac{\partial f_n}{\partial x_m}(p) \end{pmatrix}.$$

When $n = 1$, this is familiar from multivariable calculus: directional derivatives are computed by dotting with the gradient. The $n > 1$ case follows by applying this fact to each component function $f_i$.

The derivative of a composition of two functions turns out to be the composition of their derivatives:

**Proposition 7.4** (Chain rule)**.** *Suppose $\gamma : \mathbb{R}^l \to \mathbb{R}^m$ is differentiable at $x \in \mathbb{R}^l$ and $f : \mathbb{R}^m \to \mathbb{R}^n$ is differentiable at $\gamma(x)$. Then their composition is differentiable at $x$, and*

$$d(f \circ \gamma)_x = df_{\gamma(x)} \circ d\gamma_x.$$

The chain rule is an important tool, and some comments about it are in order. First, $\gamma$ need not be defined on all of $\mathbb{R}^l$, but only on a neighborhood of $x$ for the chain rule to be valid. Similarly, it is enough that $f$ be defined on a neighborhood of $\gamma(x)$.

Second, the case $l = 1$ has an important visual interpretation. In this case, $\gamma$ is a path in $\mathbb{R}^m$ and $f \circ \gamma$ is the image path in $\mathbb{R}^n$. Set $x = 0$ and $p = \gamma(0)$. The chain rule says that for all $v \in \mathbb{R}$,

$$d(f \circ \gamma)_0(v) = df_p(d\gamma_0(v)).$$

Choosing $v$ as the unit-vector $v = e_1 \in \mathbb{R}^1$ gives:

$$(f \circ \gamma)'(0) = df_p(\gamma'(0)).$$

This provides an alternative interpretation of the derivative:

**Proposition 7.5.** *If $f : \mathbb{R}^m \to \mathbb{R}^n$ is differentiable at $p \in \mathbb{R}^m$, and $v \in \mathbb{R}^m$, then $df_p(v)$ is the initial velocity vector of the composition with $f$ of any differentiable path $\gamma$ in $\mathbb{R}^m$ with $\gamma(0) = p$ and $\gamma'(0) = v$.*

This proposition says Figure 1 remains valid when the straight line is replaced by any (possibly curved) path $\gamma$ with $\gamma(0) = p$ and $\gamma'(0) = v$. This proposition is so useful, we will take it as our definition of $df_p(v)$ for the remainder of the book.

Another important consequence of the chain rule is that the derivative of an invertible function is an element of the general linear group. More precisely, suppose that $U \subset \mathbb{R}^n$ is open and $f : U \to \mathbb{R}^n$ is an invertible function from $U$ to its image $f(U)$. Suppose that $f$ is differentiable at $x \in U$ and that $f^{-1}$ is differentiable at $f(x)$. The chain rule says:

$$d(f^{-1} \circ f)_x = d(f^{-1})_{f(x)} \circ df_x.$$

On the other hand, $f^{-1} \circ f$ is the identify function whose derivative at any point is the identify map. So $d(f^{-1})_{f(x)} \circ df_x$ is the identity linear map, which means that $df_x$ is an invertible linear map (the corresponding matrix is an element of $GL_n(\mathbb{R})$).

A crucial result from analysis is the following converse:

**Theorem 7.6** (Inverse function theorem)**.** *If $f : \mathbb{R}^n \to \mathbb{R}^n$ is $C^r$ on a neighborhood of $x \in \mathbb{R}^n$ $(r \geq 1)$ and $df_x$ is an invertible linear map, then there exists a (possibly smaller) neighborhood $U$ of $x$ such that $V = f(U)$ is a neighborhood of $f(x)$, and $f : U \to V$ is invertible with $C^r$ inverse.*

The inverse function theorem is quite remarkable. It reduces the seemingly difficult problem of deciding whether $f$ is locally invertible near $x$ to the computationally simple task of checking whether the determinant of the linear map $df_x$ is non-zero! The proof is non-trivial, but the theorem is believable, since $f(y) \approx f(x) + df_x(y - x)$ is a first-order approximation of $f$ near $x$. The theorem says that if this first-order approximation is bijective, then $f$ is bijective near $x$.

## 2. Proof of part (1) of Theorem 7.1

Let $G \subset GL_n(\mathbb{K})$ be a matrix group with Lie algebra $\mathfrak{g}$. Part (1) of Theorem 7.1 says that for all $X \in \mathfrak{g}$, $e^X \in G$. We verified this for several groups. Another reason to expect the theorem to be true comes from the following idea. The tangent space to $G$ at $I$ is $\mathfrak{g}$, and the tangent space at $a \in G$ is

$$T_aG = a \cdot \mathfrak{g} = \{a \cdot Y \mid Y \in \mathfrak{g}\}$$

(by Exercise 5.10). Fix a vector $X \in \mathfrak{g}$. Consider the vector field $\mathcal{V}$ on $M_n(\mathbb{K})$ whose value at $a \in M_n(\mathbb{K})$ is $\mathcal{V}(a) = a \cdot X$. At points of $G$, this vector field is tangent to $G$. The path $\gamma(t) = e^{tX}$ is an integral curve of $\mathcal{V}$, because $\gamma'(t) = \gamma(t) \cdot X$ (by Proposition 6.10). Since $\gamma(0) = I \in G$, we expect $\gamma(t)$ to remain in $G$.

It would be nice to know that $G$ is a manifold, since an integral curve of a smooth vector field on Euclidean space which at points of a manifold $M$ is tangent to $M$ must remain on $M$ if it begins on $M$. But we're getting ahead of ourselves, since we haven't defined manifold, and we will need Theorem 7.1 in order to prove that matrix groups are manifolds. To avoid circular reasoning, we must abandon the argument, although the following proof (from [**11**]) does reflect some of its essence.

**Proof of part (1) of Theorem 7.1.** Let $\{X_1, ..., X_k\}$ be a basis of $\mathfrak{g}$. For each $i = 1, ..., k$ choose a differentiable path $\alpha_i : (-\epsilon, \epsilon) \to G$ with $\alpha_i(0) = I$ and $\alpha_i'(0) = X_i$. Define

$$F_{\mathfrak{g}} : (\text{neighborhood of } 0 \text{ in } \mathfrak{g}) \to G$$

as follows:

$$F_{\mathfrak{g}}(c_1X_1 + \cdots + c_kX_k) = \alpha_1(c_1) \cdot \alpha_2(c_2) \cdots \alpha_k(c_k).$$

Notice that $F_{\mathfrak{g}}(0) = I$, and $d(F_{\mathfrak{g}})_0$ is the identify function:

$$d(F_{\mathfrak{g}})_0(X) = X \text{ for all } X \in \mathfrak{g},$$

as is easily verified on basis elements.

Choose a subspace $\mathfrak{p} \subset M_n(\mathbb{K})$ that is complementary to $\mathfrak{g}$, which means completing the set $\{X_1, ..., X_k\}$ to a basis of all of $M_n(\mathbb{K})$ and defining $\mathfrak{p}$ as the span of the added basis elements. So $M_n(\mathbb{K}) = \mathfrak{g} \times \mathfrak{p}$.

Choose a function $F_{\mathfrak{p}} : \mathfrak{p} \to M_n(\mathbb{K})$ with $F_{\mathfrak{p}}(0) = I$ and with $d(F_{\mathfrak{p}})_0(V) = V$ for all $V \in \mathfrak{p}$. For example, $F_{\mathfrak{p}}(V) = I + V$ works. Next define the function

$$F : (\text{neighborhood of } 0 \text{ in } \mathfrak{g} \times \mathfrak{p} = M_n(\mathbb{K})) \to M_n(\mathbb{K})$$

by the rule $F(X+Y) = F_{\mathfrak{g}}(X)\cdot F_{\mathfrak{p}}(Y)$ for all $X \in \mathfrak{g}$ and $Y \in \mathfrak{p}$. Notice that $F(0) = I$ and $dF_0$ is the identity function: $dF_0(X+Y) = X+Y$.

By the inverse function theorem, $F$ has an inverse function defined on a neighborhood of $I$ in $M_n(\mathbb{K})$. Express the inverse as follows for matrices $a$ in this neighborhood:

$$F^{-1}(a) = u(a) + v(a) \in \mathfrak{g} \times \mathfrak{p}.$$

By definition, $u(F(X+Y)) = X$ and $v(F(X+Y)) = Y$ for all $X \in \mathfrak{g}$ and $Y \in \mathfrak{p}$ near 0. The important thing is that $v$ tests whether an element $a \in M_n(\mathbb{K})$ near $I$ lies in $G$:

$$v(a) = 0 \implies a \in G.$$

Let $X \in \mathfrak{g}$ and define $a(t) = e^{tX}$. We wish to prove that $a(t) \in G$ for small $t$ by showing that $v(a(t)) = 0$. Since $v(a(0)) = 0$, it will suffice to prove that $\frac{d}{dt}v(a(t)) = 0$ for small $t$. Since

$$\frac{d}{dt}v(a(t)) = dv_{a(t)}(a'(t)) = dv_{a(t)}(X \cdot a(t)),$$

the result will follow from the following lemma:

**Lemma 7.7.** *For all $a \in M_n(\mathbb{K})$ near $I$ and all $X \in \mathfrak{g}$, $dv_a(X\cdot a) = 0$.*

**Proof.** Express $a$ as:

$$a = F(Z+Y) = F_{\mathfrak{g}}(Z) \cdot F_{\mathfrak{p}}(Y),$$

where $Z \in \mathfrak{g}$ and $Y \in \mathfrak{p}$. For all $W \in \mathfrak{g}$, and for sufficiently small $t$,

$$v(F_{\mathfrak{g}}(Z+tW) \cdot F_{\mathfrak{p}}(Y)) = Y,$$

which means that $v$ is not changing at $a$ in these directions:

$$\begin{aligned} 0 &= \left.\frac{d}{dt}\right|_{t=0} v(F_{\mathfrak{g}}(Z+tW) \cdot F_{\mathfrak{p}}(Y)) \\ &= dv_a((d(F_{\mathfrak{g}})_Z(W)) \cdot F_{\mathfrak{p}}(Y)) \\ &= dv_a((d(F_{\mathfrak{g}})_Z(W)) \cdot F_{\mathfrak{g}}(Z)^{-1} \cdot a) \\ &= dv_a(X \cdot a), \end{aligned}$$

where $X = (d(F_{\mathfrak{g}})_Z(W)) \cdot F_{\mathfrak{g}}(Z)^{-1}$. It remains to prove that $X$ is an arbitrary element of $\mathfrak{g}$. First, $X \in \mathfrak{g}$ because it is the initial tangent vector of the following path in $G$:

$$t \mapsto F_{\mathfrak{g}}(Z + tW) \cdot F_{\mathfrak{g}}(Z)^{-1}.$$

Second, $X$ is arbitrary because the linear map from $\mathfrak{g} \to \mathfrak{g}$ which sends

$$W \mapsto (d(F_{\mathfrak{g}})_Z(W)) \cdot F_{\mathfrak{g}}(Z)^{-1}$$

is the identity map when $Z = 0$, and so by continuity has determinant close to 1, and is therefore an isomorphism, when $Z$ is close to 0. In other words, $W$ can be chosen so that $X$ is any element of $\mathfrak{g}$. □

The lemma completes our proof that if $X \in \mathfrak{g}$, then $e^{tX} \in G$ for small $t$, say for $t \in (-\epsilon, \epsilon)$. The result can be extended by observing that for all $t \in (-\epsilon, \epsilon)$ and all positive integers $N$,

$$e^{NtX} = e^{tX + tX + \cdots + tX} = e^{tX} \cdot e^{tX} \cdots e^{tX} \in G.$$

This verifies that $e^{tX} \in G$ for all $t \in \mathbb{R}$, which completes the proof! □

## 3. Proof of part (2) of Theorem 7.1

It can be shown that any power series gives a smooth function on the set of matrices with norm less than its radius of convergence. In particular:

**Proposition 7.8.** $\exp : M_n(\mathbb{K}) \to M_n(\mathbb{K})$ *is smooth.*

This fact allows us to verify part (2) of Theorem 7.1 in the special case $G = GL_n(\mathbb{K})$. Remember that $B_r = \{W \in M_n(\mathbb{K}) \mid |W| < r\}$.

**Lemma 7.9.** *For sufficiently small $r > 0$, $V = \exp(B_r)$ is a neighborhood of $I$ in $GL_n(\mathbb{K})$, and $\exp : B_r \to V$ is a homeomorphism (which is smooth and has smooth inverse).*

**Proof.** For all $X \in M_n(\mathbb{K})$, $d(\exp)_0(X)$ is the initial tangent vector to the path $t \mapsto e^{tX}$, namely $X$. In other words, $d(\exp)_0$ is the identity map. The result now follows from the inverse function theorem, together with the observation that a sufficiently small neighborhood of $I$ in $M_n(\mathbb{K})$ must lie in $GL_n(\mathbb{K})$. □

The inverse of exp is denoted "log"; it is a smooth function defined on a neighborhood of $I$ in $GL_n(\mathbb{K})$. Although we will not require this fact, it is not hard to prove that $\log(A)$ equals the familiar power series for log evaluated on $A$:

$$\log(A) = (A-I) - (1/2)(A-I)^2 + (1/3)(A-I)^3 - (1/4)(A-I)^4 + \cdots .$$

Now let $G \subset GL_n(\mathbb{K})$ be a matrix group with Lie algebra $\mathfrak{g}$. Part 2 of Proposition 7.1 says that for sufficiently small $r > 0$, $\exp(B_r \cap \mathfrak{g})$ is a neighborhood of $I$ in $G$. Lemma 7.9 handled the case where $G$ is all of $GL_n(\mathbb{K})$. Generalizing to arbitrary $G$ is not as obvious as it might at first seem. In fact, the proposition can be false for a subgroup $G \subset GL_n(\mathbb{K})$ that is not closed, as the next example illustrates.

**Example 7.10.** *Let $\lambda \in \mathbb{R}$ be an irrational number, and define*

$$G = \left\{ g_t = \begin{pmatrix} e^{t\mathbf{i}} & 0 \\ 0 & e^{\lambda t\mathbf{i}} \end{pmatrix} \middle| t \in \mathbb{R} \right\} \subset GL_2(\mathbb{C}).$$

*The Lie algebra of $G$ is the span of $W = \begin{pmatrix} \mathbf{i} & 0 \\ 0 & \lambda\mathbf{i} \end{pmatrix}$, and $e^{tW} = g_t$ for all $t \in \mathbb{R}$. For $0 < r < \infty$, notice that*

$$\exp(\{tW \mid t \in (-r,r)\}) = \{g_t \mid t \in (-r,r)\}$$

*is not a neighborhood of $I$ in $G$. Any neighborhood of $I$ in $G$ contains points of the form $g_{2\pi n}$ for arbitrarily large integers $n$; compare with Exercise 4.24.*

We require the following important lemma:

**Lemma 7.11.** *Let $G \subset GL_n(\mathbb{K})$ be a matrix group with Lie algebra $\mathfrak{g}$. In Lemma 7.9, $r > 0$ can be chosen such that additionally:*

$$\exp(B_r \cap \mathfrak{g}) = \exp(B_r) \cap G.$$

For any $r$, $\exp(B_r \cap \mathfrak{g}) \subset \exp(B_r) \cap G$, so the real content of this lemma is that the other inclusion holds for sufficiently small $r$. The lemma is false for certain non-closed subgroups of $GL_n(\mathbb{K})$, including the one in Example 7.10. The essential problem is this: there are elements of $G$ (namely $g_{2\pi n}$ for certain large $n$) that are arbitrarily close to $I$, so they are exponential images of arbitrarily short vectors in $M_n(\mathbb{K})$, but they are exponential images only of very long vectors in $\mathfrak{g}$.

**Proof of Lemma 7.11.** Choose a subspace $\mathfrak{p} \subset M_n(\mathbb{K})$ which is complementary to $\mathfrak{g}$, as in the proof of part (1) of Theorem 7.1, so $M_n(\mathbb{K}) = \mathfrak{g} \times \mathfrak{p}$. Define the function $\Phi : \mathfrak{g} \times \mathfrak{p} \to M_n(\mathbb{K})$ so that $\Phi(X+Y) = e^X e^Y$ for all $X \in \mathfrak{g}$ and $Y \in \mathfrak{p}$. Notice that $\Phi$ agrees with exp on $\mathfrak{g}$. The functions $\Phi$ and exp are also similar in that the derivative of each at 0 is the identity. In particular, $\Phi$ is locally invertible by the inverse function theorem.

Assume the lemma is false. Then there must be a sequence of non-zero vectors $\{A_1, A_2, ...\}$ in $M_n(\mathbb{K})$ with $|A_i| \to 0$ such that $A_i \notin \mathfrak{g}$ and $\Phi(A_i) \in G$ for all $i$. Write $A_i = X_i + Y_i$, where $X_i \in \mathfrak{g}$ and $0 \neq Y_i \in \mathfrak{p}$. For all $i$, let $g_i = \Phi(A_i) = e^{X_i} e^{Y_i} \in G$. Notice that $e^{Y_i} = (e^{X_i})^{-1} g_i \in G$.

By compactness of the sphere of unit-length vectors in $\mathfrak{p}$, the sequence $\{\frac{Y_1}{|Y_1|}, \frac{Y_2}{|Y_2|}, ...\}$ must subconverge to some unit-length vector $Y \in \mathfrak{p}$ (by Proposition 4.23). For notational convenience, re-choose the $A_i$'s above so that the sequence converges rather than subconverges to $Y$.

Let $t \in \mathbb{R}$. Since $|Y_i| \to 0$, it is possible to choose a sequence of positive integers $n_i$ such that $n_i Y_i \to tY$. Since $e^{n_i Y_i} = (e^{Y_i})^{n_i} \in G$, and since $G$ is closed in $GL_n(\mathbb{K})$, it follows that $e^{tY} \in G$. In summary, $e^{tY} \in G$ for all $t \in \mathbb{R}$, which is impossible since $Y \notin \mathfrak{g}$. $\square$

**Proof of part (2) of Theorem 7.1.** Pick $r > 0$ as in Lemma 7.11. Then $V = \exp(B_r \cap \mathfrak{g})$ is a neighborhood of $I$ in $G$ because it equals the set $\exp(B_r) \cap G$, and $\exp(B_r)$ is open in $M_n(\mathbb{K})$ by Lemma 7.9. The restriction $\exp : B_r \cap \mathfrak{g} \to V$ is continuous. Its inverse function $\log : V \to B_r \cap \mathfrak{g}$ is continuous because it is a restriction of the continuous function $\log : \exp(B_r) \to B_r$. $\square$

In the previous proof, $\exp : B_r \cap \mathfrak{g} \to V$ is not only continuous, it is smooth. Its inverse $\log : V \to B_r \cap \mathfrak{g}$ is also better than continuous; it is the restriction to $V$ of the smooth function log.

## 4. Manifolds

In this section, we define manifolds and prove that matrix groups are manifolds.

Let $X \subset \mathbb{R}^m$ be any subset, and let $f : X \to \mathbb{R}^n$ be a function. If $X$ is open, it makes sense to ask whether $f$ is smooth. If $X$ is not open, then the partial derivatives of $f$ at $p \in X$ might not make sense, because $f$ need not be defined near $p$ in all coordinate directions. We will call $f$ smooth if it locally extends to a smooth function on $\mathbb{R}^m$:

**Definition 7.12.** *If $X \subset \mathbb{R}^m$, then $f : X \to \mathbb{R}^n$ is called **smooth** if for all $p \in X$, there exists a neighborhood, $U$, of $p$ in $\mathbb{R}^m$ and a smooth function $\tilde{f} : U \to \mathbb{R}^n$ that agrees with $f$ on $X \cap U$.*

This extended notion of smoothness allows us to define an important type of equivalence for subsets of Euclidean space:

**Definition 7.13.** *$X \subset \mathbb{R}^{m_1}$ and $Y \subset \mathbb{R}^{m_2}$ are called **diffeomorphic** if there exists a smooth bijective function $f : X \to Y$ whose inverse is also smooth. In this case, $f$ is called a **diffeomorphism**.*

From the discussion after its proof, it is clear that the word "homeomorphism" can be replaced by the word "diffeomorphism" in part (2) of Theorem 7.1.

A diffeomorphism is a homeomorphism that is smooth and has a smooth inverse. Figure 2 shows two sets that are homeomorphic but are not diffeomorphic, because no homeomorphism between them could be smooth at the cone point.

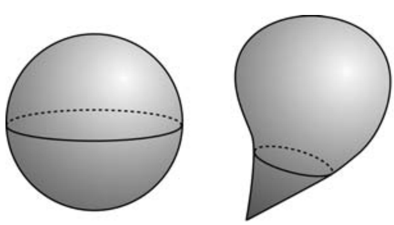


**Figure 2.** Homeomorphic but not diffeomorphic.

A manifold is a set that is locally diffeomorphic to Euclidean space:

**Definition 7.14.** *A subset $X \subset \mathbb{R}^m$ is called an* ***(embedded) manifold*** *of dimension $d$ if for all $p \in X$ there exists a neighborhood, $V$, of $p$ in $X$ that is diffeomorphic to an open set $U \subset \mathbb{R}^d$.*

We will omit the word "embedded" and just call it a "manifold" for now, until we learn in Chapter 10 about a more general type of manifold. In Figure 2, the round sphere is a 2-dimensional manifold in $\mathbb{R}^3$. A sufficiently small and nearsighted bug living on this sphere would think it lived on $\mathbb{R}^2$. But the pointed sphere is not a manifold, since a bug living at the cone point, no matter how nearsighted, could distinguish its home from $\mathbb{R}^2$.

Bugs are fine, but to rigorously prove that a set $X$ is a manifold, one must construct a **parametrization** at every $p \in X$, meaning a diffeomorphism, $\varphi$, from an open set $U \subset \mathbb{R}^d$ to a neighborhood, $V$, of $p$ in $X$, as in Figure 3.

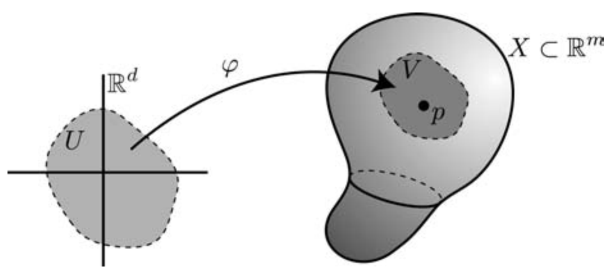



**Figure 3.** A parametrization of $X$ at $p$.

For practice, we will prove $S^2 = \{(x, y, x) \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1\}$ is a 2-dimensional manifold.

**Proposition 7.15.** *$S^2 \subset \mathbb{R}^3$ is a 2-dimensional manifold.*

**Proof.** The upper hemisphere

$$V = \{(x, y, z) \in S^2 \mid z > 0\}$$

is a neighborhood of $(0, 0, 1)$ in $S^2$. Define

$$U = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\},$$

and define $\varphi : U \to V$ as $\varphi(x,y) = (x, y, \sqrt{1-x^2-y^2})$. Then $\varphi$ is smooth and bijective. The inverse $\varphi^{-1} : V \to U$ has the formula $\varphi^{-1}(x,y,z) = (x,y)$. By Definition 7.12, $\varphi^{-1}$ is smooth because it extends to the smooth function with this same formula defined on the open set $\{(x,y,z) \in \mathbb{R}^3 \mid z > 0\}$, or even on all of $\mathbb{R}^3$.

For arbitrary $p \in S^2$, the function $(R_A) \circ \varphi : U \to R_A(V)$ is a parametrization at $p$, assuming $A \in SO(3)$ is any matrix for which $R_A(0,0,1) = p$. □

Before proving that all matrix groups are manifolds, we give a simple example:

**Claim 7.16.** *The matrix group*

$$T = \{diag(e^{\mathbf{i}\theta}, e^{\mathbf{i}\phi}) \mid \theta, \phi \in [0, 2\pi)\} \subset GL_2(\mathbb{C})$$

*is a 2-dimensional manifold in* $M_2(\mathbb{C}) \cong \mathbb{C}^4 \cong \mathbb{R}^8$.

**Proof.** Making the identification $M_2(\mathbb{C}) \cong \mathbb{R}^8$ explicit, we write:

$$T = \{(\cos\theta, \sin\theta, 0, 0, 0, 0, \cos\phi, \sin\phi) \mid \theta, \phi \in [0, 2\pi)\} \subset \mathbb{R}^8.$$

The identity element of $T$ is $p = (1,0,0,0,0,0,1,0)$. To describe a parametrization of $T$ at $p$, let $U = \{(\theta, \phi) \in \mathbb{R}^2 \mid -\pi/2 < \theta, \phi < \pi/2\}$ and define $\varphi : U \to T$ as $(\theta, \phi) \mapsto (\cos\theta, \sin\theta, 0, 0, 0, 0, \cos\phi, \sin\phi)$. This parametrization is clearly smooth and is bijective onto its image $V = \varphi(U)$. The inverse $\varphi^{-1} : V \to U$ is also smooth because it extends to the smooth function from an open set in $\mathbb{R}^8$ to $U$ defined as follows:

$$(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) \mapsto (\arctan(x_2/x_1), \arctan(x_8/x_7)).$$

A parametrization at an arbitrary point of $T$ is defined similarly. □

In the previous claim, notice that $T$ is diffeomorphic to the manifold in $\mathbb{R}^4$ obtained by removing the four irrelevant components of $\mathbb{R}^8$. In fact, $T$ is diffeomorphic to a manifold in $\mathbb{R}^3$, namely, the **torus of revolution** obtained by revolving about the $z$-axis a circle in the $yz$-plane. This assertion comes from observing that each point on this torus of revolution is described by a pair of angles: $\theta$ describes a point on the circle in the $yz$-plane, and $\phi$ describes how far that point rotates about the $z$-axis. (See Figure 4.)

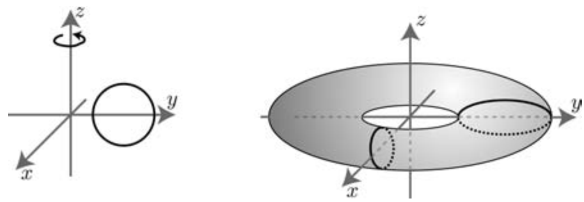



**Figure 4.** A torus of revolution in $\mathbb{R}^3$.

**Theorem 7.17.** *Any matrix group of dimension $d$ is a manifold of dimension $d$.*

**Proof.** Let $G \subset GL_n(\mathbb{K})$ be a matrix group of dimension $d$ with Lie algebra $\mathfrak{g}$. Choose $r > 0$ as in Theorem 7.1. Then $V = \exp(B_r \cap \mathfrak{g})$ is a neighborhood of $I$ in $G$, and the restriction $\exp : B_r \cap \mathfrak{g} \to V$ is a parametrization at $I$. Here we are implicitly identifying $\mathfrak{g}$ with $\mathbb{R}^d$ by choosing a basis.

Next let $g \in G$ be arbitrary. Define $\mathcal{L}_g : M_n(\mathbb{K}) \to M_n(\mathbb{K})$ as

$$\mathcal{L}_g(A) = g \cdot A.$$

Notice that $\mathcal{L}_g$ restricts to a diffeomorphism from $G$ to $G$. So $\mathcal{L}_g(V)$ is a neighborhood of $g$ in $G$, and $(\mathcal{L}_g \circ \exp) : B_r \cap \mathfrak{g} \to \mathcal{L}_g(V)$ is a parametrization at $g$. $\square$

## 5. More about manifolds

In this section, we prove that each tangent space to a manifold is a subspace, and we define the "derivative" of a smooth function $f : X_1 \to X_2$ between manifolds. The derivative of $f$ at $p \in X_1$ will be denoted $df_p$ and is a linear map from $T_pX_1$ to $T_{f(p)}X_2$. In future sections, our primary applications will be when $f$ is a smooth homomorphism between matrix groups, in which case $df_I$ is a linear map between their Lie algebras.

Remember that in Definition 5.1, the tangent space to a subset $X \subset \mathbb{R}^m$ at $p \in X$ was defined as:

$$T_pX = \{\gamma'(0) \mid \gamma : (-\epsilon, \epsilon) \to X \text{ is differentiable with } \gamma(0) = p\}.$$

**Proposition 7.18.** *If $X \subset \mathbb{R}^m$ is a $d$-dimensional manifold, then for each $p \in X$, $T_pX$ is a $d$-dimensional subspace of $\mathbb{R}^m$.*

**Proof.** To prove the proposition, we will present a more technical definition of $T_pX$ and then prove that the two definitions are equivalent. Let $\varphi : U \subset \mathbb{R}^d \to V \subset X$ be a parametrization at $p$. Assume for simplicity that $0 \in U$ and $\varphi(0) = p$. Define

$$T_pX = d\varphi_0(\mathbb{R}^d).$$

This makes sense if $\varphi$ is regarded as a function from $U \subset \mathbb{R}^d$ to $\mathbb{R}^m$. Clearly $T_pX$ is a subspace, since it's the image of a linear map. The two definitions of $T_pX$ agree because differentiable paths through $p$ in $X$ are exactly the images under $\varphi$ of differentiable paths in $U$ through 0. In particular, this agreement shows that the technical definition of $T_pX$ is well-defined; it does not depend on the choice of parametrization, $\varphi$.

To prove that the dimension of $T_pX$ equals $d$, we must confirm that $d\varphi_0$ has rank $d$. According to Definition 7.12, $\varphi^{-1} : V \to U$ near $p$ extends to a smooth function from a neighborhood of $p$ in $\mathbb{R}^m$ to $\mathbb{R}^d$. We will abuse notation by denoting this extension also as $\varphi^{-1}$. The chain rule says that:

$$d(\varphi^{-1} \circ \varphi)_0 = d(\varphi^{-1})_p \circ d\varphi_0.$$

On the other hand, $\varphi^{-1} \circ \varphi$ is the identity function, so its derivative at 0 is the identity linear transformation. Thus, $d(\varphi^{-1})_p \circ d\varphi_0$ is the identity linear transformation of $\mathbb{R}^d$, which implies that $d\varphi_0$ has maximal rank. □

Next we define the derivative of a function between manifolds. The definition is analogous to Proposition 7.5 and is pictured in Figure 5.

**Definition 7.19.** *Let $f : X_1 \to X_2$ be a smooth function between manifolds $X_1 \subset \mathbb{R}^{m_1}$ and $X_2 \subset \mathbb{R}^{m_2}$. Let $p \in X_1$. If $v \in T_pX_1$, then $df_p(v) \in T_{f(p)}X_2$ denotes the initial velocity vector, $(f \circ \gamma)'(0)$, of the composition with $f$ of any differentiable path $\gamma$ in $X_1$ with $\gamma(0) = p$ and $\gamma'(0) = v$.*

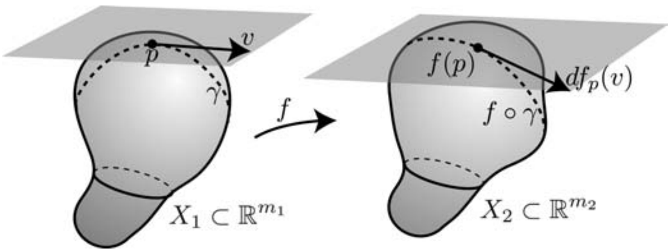



**Figure 5.** $df_p(v)$ means the initial velocity vector of the composition with $f$ of any path $\gamma$ through $p$ in the direction of $v$.

**Proposition 7.20.** *Under the hypotheses of Definition* 7.19*, the map* $v \mapsto df_p(v)$ *is a well-defined linear transformation from* $T_pX_1$ *to* $T_{f(p)}X_2$.

Here "well-defined" means independent of the choice of $\gamma$.

**Proof.** The smoothness of $f$ means that there is a neighborhood, $\mathcal{O}$, of $p$ in $\mathbb{R}^{m_1}$ and a smooth function $\tilde{f} : \mathcal{O} \to \mathbb{R}^{m_2}$ that agrees with $f$ on $\mathcal{O} \cap X_1$. For any $v \in \mathbb{R}^{m_1}$, Proposition 7.5 says that $d\tilde{f}_p(v) = (\tilde{f} \circ \gamma)'(0)$, where $\gamma$ is any regular curve in $\mathbb{R}^{m_1}$ with $\gamma(0) = p$ and $\gamma'(0) = v$. If $v \in T_pX_1$, then such a regular curve $\gamma$ can be chosen to lie in $X_1$, in which case: $d\tilde{f}_p(v) = (\tilde{f} \circ \gamma)'(0) = (f \circ \gamma)'(0) = df_p(v)$. In summary, $df_p$ equals the restriction to the domain $T_pX_1$ of the well-defined linear transformation $d\tilde{f}_p : \mathbb{R}^{m_1} \to \mathbb{R}^{m_2}$. $\square$

The two most important facts about derivatives of functions between Euclidean spaces generalize to functions between manifolds:

**Proposition 7.21** (Chain rule for manifolds)**.** *Suppose* $f : X_1 \to X_2$ *and* $g : X_2 \to X_3$ *are smooth functions between manifolds. Then so is their composition, and for all* $p \in X_1$,

$$d(g \circ f)_p = dg_{f(p)} \circ df_p.$$

**Theorem 7.22** (Inverse function theorem for manifolds)**.** *Suppose* $f : X_1 \to X_2$ *is a smooth function between manifolds, and* $p \in X_1$. *If* $df_p : T_pX_1 \to T_{f(p)}X_2$ *is an invertible linear transformation, then there exists a neighborhood,* $U$*, of* $p$ *in* $X_1$ *such that* $V = f(U)$ *is* *a neighborhood of $f(p)$ in $X_2$, and the restriction $f : U \to V$ is a diffeomorphism.*

**Proof.** Since $df_p$ is invertible, $X_1$ and $X_2$ must have the same dimension, which we denote as $d$. Let $\varphi_1 : U_1 \subset \mathbb{R}^d \to V_1 \subset X_1$ be a parametrization at $p$ with $\varphi_1(0) = p$. Let $\varphi_2 : U_2 \subset \mathbb{R}^d \to V_2 \subset X_2$ be a parametrization at $f(p)$ with $\varphi_2(0) = f(p)$. The fact that $f$ is smooth implies that

$$\phi = \varphi_2^{-1} \circ f \circ \varphi_1 : U_1 \to U_2$$

is smooth (one may have to shrink $U_1$ to a smaller neighborhood of 0 in $\mathbb{R}^d$ in order for $\phi$ to be defined on $U_1$); see Figure 6.

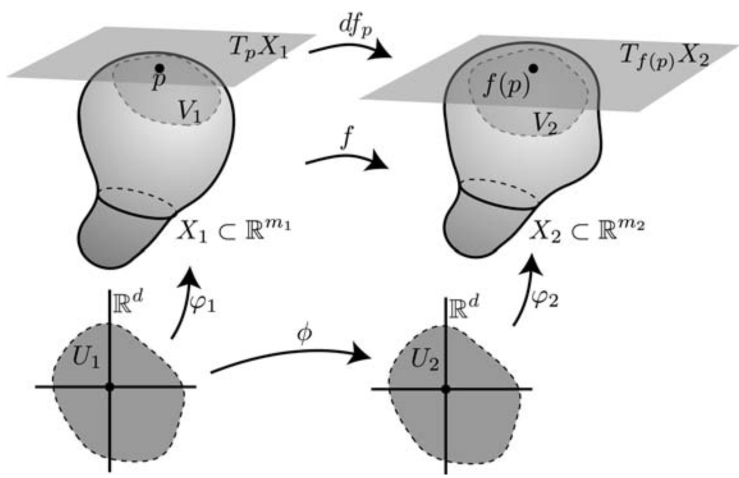



**Figure 6.** $\phi = \varphi_2^{-1} \circ f \circ \varphi_1 : U_1 \to U_2$.

By the chain rule (applied twice):

$$d\phi_0 = d(\varphi_2^{-1})_{f(p)} \circ df_p \circ d(\varphi_1)_0 : T_pX_1 \to T_{f(p)}X_2.$$

Thus, $d\phi_0$ is the composition of three invertible linear transformations and is therefore invertible. The inverse function theorem (7.6) implies that, after possibly further shrinking the neighborhoods in question, $\phi : U_1 \to U_2$ is a diffeomorphism. Therefore,

$$f = \varphi_2 \circ \phi \circ (\varphi_1)^{-1} : V_1 \to V_2$$

is the composition of three diffeomorphisms, so it is a diffeomorphism. □

We will consider matrix groups $G_1$ and $G_2$ equivalent if there exists a group-isomorphism $f : G_1 \to G_2$ that is also a diffeomorphism (so they simultaneously look the same as groups and as manifolds). In this case $G_1$ and $G_2$ will be called **smoothly isomorphic**. Smoothly isomorphic matrix groups have the same dimension (according to Exercise 7.10).

There is a non-trivial theorem that any continuous homomorphism between matrix groups is smooth. Therefore, "continuously isomorphic" is the same as "smoothly isomorphic"; see [**12**] or [**13**] for a proof. On the other hand, it is possible for a homomorphism between matrix groups to be discontinuous. For example, the additive group $(\mathbb{R}, +)$ can be considered a matrix group, because it is isomorphic to $\text{Trans}(\mathbb{R}^1)$. There are many discontinuous isomorphisms $f : (\mathbb{R}, +) \to (\mathbb{R}, +)$. In fact, any bijection of a basis for $\mathbb{R}$ (regarded as a vector space over $\mathbb{Q}$) extends linearly to an isomorphism.

## 6. Exercises

**Ex. 7.1.** The strategy used in Proposition 7.15 to prove that $S^2$ is a manifold requires at least six parametrizations (respectively covering the hemispheres $z > 0$, $z < 0$, $x > 0$, $x < 0$, $y > 0$ and $y < 0$). Another common strategy requires only two parametrizations. For this, the **stereographic projection** function, $f : S^2 - \{(0, 0, 1)\} \to \mathbb{R}^2$, is defined so that $f(p)$ equals the $xy$-coordinates of the intersection of the plane $z = -1$ with the line containing $(0, 0, 1)$ and $p$; see Figure 7.

(1) Use similar triangles to verify that:
$$f(x, y, z) = \frac{2}{1 - z}(x, y).$$
(2) Find a formula for $f^{-1} : \mathbb{R}^2 \to S^2 - \{(0, 0, 1)\}$.
(3) Find a formula for the function $g : S^2 - \{(0, 0, -1)\} \to \mathbb{R}^2$, defined so that $g(p)$ equals the $xy$-coordinates of the intersection of the plane $z = 1$ with the line containing $(0, 0, -1)$ and $p$. Also find a formula for $g^{-1} : \mathbb{R}^2 \to S^2 - \{(0, 0, -1)\}$.

Notice that the parametrizations $f^{-1}$ and $g^{-1}$ together cover $S^2$, which verifies that it is a manifold.

(4) Find an explicit formula for the composition

$$g \circ f^{-1} : \mathbb{R}^2 - \{(0,0)\} \to \mathbb{R}^2 - \{(0,0)\}.$$

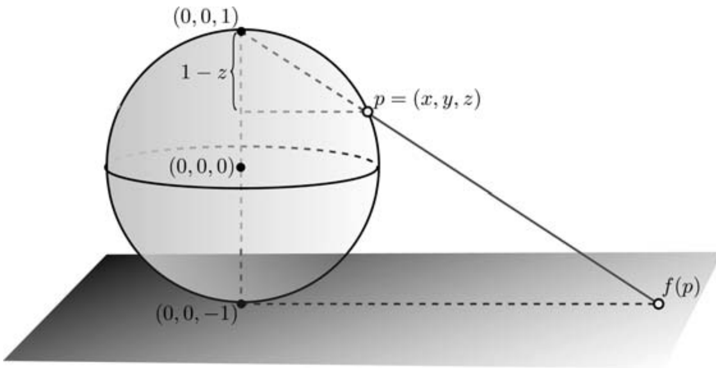



**Figure 7.** Stereographic projection shadows $S^2$ onto the plane $z = -1$.

**Ex. 7.2.** Prove that $S^n \subset \mathbb{R}^{n+1}$ is an $n$-dimensional manifold.

**Ex. 7.3.** Prove that the cone $\{(x,y,z) \in \mathbb{R}^3 \mid z = \sqrt{x^2+y^2}\} \subset \mathbb{R}^3$ is not a manifold.

**Ex. 7.4.** If $X_1 \subset \mathbb{R}^{m_1}$ and $X_2 \subset \mathbb{R}^{m_2}$ are manifolds whose dimensions are $d_1$ and $d_2$, prove that

$$X_1 \times X_2 = \{(p_1,p_2) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \cong \mathbb{R}^{m_1+m_2} \mid p_1 \in X_1, p_2 \in X_2\}$$

is a $d_1 + d_2$ dimensional manifold.

**Ex. 7.5.** Is the group $G$ in Example 7.10 a manifold?

**Ex. 7.6.** Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a *linear* function. For any $p \in \mathbb{R}^m$, show that $df_p = f$. In other words, the derivative of a linear function is itself.

**Ex. 7.7.** Let $G$ be a (not necessarily path-connected) matrix group. Define the **identity component**, $G_0$, of $G$ as:

$$\{g \in G \mid \exists \text{ continuous } \gamma : [0,1] \to G \text{ with } \gamma(0) = I \text{ and } \gamma(1) = g\}.$$

Prove that $G_0$ is a matrix group (don't forget to prove $G_0$ is closed). Prove that $G_0$ is a normal subgroup of $G$.

**Ex. 7.8.** Prove that there exists a neighborhood of $I$ in $GL_n(\mathbb{K})$ that does not contain any subgroup of $GL_n(\mathbb{K})$ other than $\{I\}$.

**Ex. 7.9.** Prove the chain rule for manifolds (Proposition 7.21).

**Ex. 7.10.** If $f : X_1 \to X_2$ is a diffeomorphism, and $p \in X_1$, prove that $df_p : T_pX_1 \to T_{f(p)}X_2$ is a linear isomorphism. In particular, $X_1$ and $X_2$ have the same dimension.

**Ex. 7.11.** Let $X$ be a manifold. If $X$ contains no clopen subsets other than itself and the empty set, prove that $X$ is path-connected (this is a converse of Proposition 4.17 for manifolds).

**Ex. 7.12.** Let $X \subset \mathbb{R}^m$ be a manifold. Define the **tangent bundle** of $X$ as:

$$TX = \{(p, v) \in \mathbb{R}^m \times \mathbb{R}^m \cong \mathbb{R}^{2m} \mid p \in X \text{ and } v \in T_pX\}.$$

Prove that $TX$ is a manifold of dimension twice the dimension of $X$.

**Ex. 7.13.** If $X \subset \mathbb{R}^m$ is a manifold, define the **unit tangent bundle** of $X$ as:

$$T^1X = \{(p, v) \in \mathbb{R}^m \times \mathbb{R}^m \cong \mathbb{R}^{2m} \mid p \in X \text{ and } v \in T_pX \text{ and } |v| = 1\}.$$

Prove that $T^1X$ is a manifold of dimension one less than the dimension of $TX$.

**Ex. 7.14.** Describe a diffeomorphism between $SO(3)$ and $T^1S^2$ (compare to Exercise 1.1).

**Ex. 7.15.** Let $f : \mathbb{R}^n \to \mathbb{R}^n$ ($n \geq 2$) be a diffeomorphism that sends translated lines to translated lines, in the sense of Exercise 3.11. Prove that $f$ has the formula $f(X) = R_A(X) + V$ for some $A \in GL_n(\mathbb{R})$ and some $V \in \mathbb{R}^n$; in other words $f$ is represented by an element of $\text{Aff}_n(\mathbb{R})$. *Hint: First prove that the matrix $df_p$ is independent of the choice of $p \in \mathbb{R}^n$.*

**Ex. 7.16.** Let $G_1$ and $G_2$ be matrix groups with Lie algebras $\mathfrak{g}_1$ and $\mathfrak{g}_2$. Let $f : G_1 \to G_2$ be a $C^1$ homomorphism. Notice that $df_I : \mathfrak{g}_1 \to \mathfrak{g}_2$. Prove that for all $v \in \mathfrak{g}_1$,

$$f(e^v) = e^{df_I(v)}.$$

In other words, a $C^1$ homomorphism is completely determined by its derivative at the identity, at least in a neighborhood of the identity. *Hint: Use Proposition* 6.17. Conclude that any $C^1$ homomorphism is smooth, at least in a neighborhood of the identity.

**Ex. 7.17.** Let $X \subset \mathbb{R}^m$ be a manifold. Let $f : X \to \mathbb{R}$ be a smooth function. Define the **graph** of $f$ as:

$$\Lambda = \{(p,t) \in \mathbb{R}^m \times \mathbb{R} \cong \mathbb{R}^{m+1} \mid p \in X \text{ and } f(p) = t\}.$$

Prove that $\Lambda$ is a manifold.

**Ex. 7.18.** Let $\varphi : Sp(1) \to O(3)$ denote the homomorphism defined in Exercise 3.16.

(1) Prove that $\varphi$ is a **local diffeomorphism**, which means there is a neighborhood of each point of its domain restricted to which $\varphi$ is a diffeomorphism onto its image.

(2) Exercise 4.27 verified that the image of $\varphi$ is contained in $SO(3)$. Prove now that the image of $\varphi$ equals $SO(3)$.

**Ex. 7.19.** Let $F : Sp(1) \times Sp(1) \to O(4)$ denote the homomorphism defined in Exercise 3.17.

(1) Prove that $F$ is a local diffeomorphism (as defined in the previous exercise).

(2) Exercise 4.28 verified that the image of $F$ is contained in $SO(4)$. Prove now that the image of $F$ equals $SO(4)$.

Chapter 8

# The Lie bracket

Since dimension is the only invariant of vector spaces, any two matrix groups of the same dimension have Lie algebras that are isomorphic as vector spaces. So how can we justify our previous assertion that the Lie algebra $\mathfrak{g}$ encodes a surprising amount of information about the matrix group $G$? In this chapter, we define the "Lie bracket" operation on $\mathfrak{g}$. For vectors $A, B \in \mathfrak{g}$, the Lie bracket $[A, B] \in \mathfrak{g}$ encodes information about the products of elements of $G$ in the directions of $A$ and $B$. Together with its Lie bracket operation, $\mathfrak{g}$ encodes information about what $G$ looks like near the identity, not just as a manifold, but also as a group.

## 1. The Lie bracket

Let $G$ be a matrix group with Lie algebra $\mathfrak{g}$. For all $g \in G$, the conjugation map $C_g : G \to G$, defined as

$$C_g(a) = gag^{-1},$$

is a smooth isomorphism. The derivative $d(C_g)_I : \mathfrak{g} \to \mathfrak{g}$ is a vector space isomorphism, which we denote as $\mathrm{Ad}_g$:

$$\mathrm{Ad}_g = d(C_g)_I.$$

To derive a simple formula for $Ad_g$, notice that any $B \in \mathfrak{g}$ can be represented as $B = b'(0)$, where $b$ is a differentiable path in $G$ with

$b(0) = I$. The product rule gives:

$$\mathrm{Ad}_g(B) = d(C_g)_I(B) = \frac{d}{dt}\Big|_{t=0} gb(t)g^{-1} = gBg^{-1}.$$

So we learn that:

$$Ad_g(B) = gBg^{-1}. \tag{8.1}$$

If we had used Equation 8.1 as our *definition* of $\mathrm{Ad}_g(B)$, it might not have been obvious that $\mathrm{Ad}_g(B) \in \mathfrak{g}$.

If all elements of $G$ commute with $g$, then $\mathrm{Ad}_g$ is the identity map on $\mathfrak{g}$. So in general, $\mathrm{Ad}_g$ measures the failure of $g$ to commute with elements of $G$ near $I$. More specifically, $Ad_g(B)$ measures the failure of $g$ to commute with elements of $G$ near $I$ in the direction of $B$. Investigating this phenomenon when $g$ is itself close to $I$ leads one to define:

**Definition 8.1.** *The **Lie bracket** of two vectors $A$ and $B$ in $\mathfrak{g}$ is:*

$$[A, B] = \frac{d}{dt}\Big|_{t=0} Ad_{a(t)}B,$$

*where $a(t)$ is any differentiable path in $G$ with $a(0) = I$ and $a'(0) = A$.*

Notice $[A, B] \in \mathfrak{g}$, since it is the initial velocity vector of a path in $\mathfrak{g}$. It measures the failure of elements of $G$ near $I$ in the direction of $A$ to commute with elements of $G$ near $I$ in the direction of $B$. The following alternative definition is easier to calculate and verifies that Definition 8.1 is independent of the choice of path $a(t)$.

**Proposition 8.2.** *For all $A, B \in \mathfrak{g}$, $[A, B] = AB - BA$.*

**Proof.** Let $a(t)$ be a differentiable path in $G$ with $a(0) = I$ and $a'(0) = A$. Using the product rule and Equation 5.1:

$$[A, B] = \frac{d}{dt}\Big|_{t=0} a(t)Ba(t)^{-1} = AB - BA.$$

□

Notice $[A, B] = 0$ if and only if $A$ and $B$ commute. The commutativity of $A$ and $B$ reflects the commutativity of elements of $G$ in the directions of $A$ and $B$. One precise way to formulate this is:

**Proposition 8.3.** *Let $A, B \in \mathfrak{g}$.*

(1) *If* $[A,B]=0$*, then* $e^{tA}$ *commutes with* $e^{sB}$ *for all* $t,s\in\mathbb{R}$*.*

(2) *If* $\exists\epsilon>0$ *such that* $e^{tA}$ *and* $e^{sB}$ *commute for* $t,s\in(-\epsilon,\epsilon)$*, then* $[A,B]=0$*.*

**Proof.** For (1), if $A$ and $B$ commute, then Proposition 6.11 gives:

$$e^{tA}e^{sB}=e^{tA+sB}=e^{sB+tA}=e^{sB}e^{tA}.$$

For (2), fix $t\in(-\epsilon,\epsilon)$, and notice that

$$Ad_{(e^{tA})}B=e^{tA}Be^{-tA}=\frac{d}{ds}\Big|_{s=0}e^{tA}e^{sB}e^{-tA}=\frac{d}{ds}\Big|_{s=0}e^{sB}=B,$$

so $[A,B]=\frac{d}{dt}\big|_{t=0}Ad_{(e^{tA})}B=0$. □

The following properties of the Lie bracket follow immediately from Proposition 8.2:

**Proposition 8.4.** *For all* $A,A_1,A_2,B,B_1,B_2,C\in\mathfrak{g}$ *and* $\lambda_1,\lambda_2\in\mathbb{R}$*,*

(1) $[\lambda_1A_1+\lambda_2A_2,B]=\lambda_1[A_1,B]+\lambda_2[A_2,B]$.

(2) $[A,\lambda_1B_1+\lambda_2B_2]=\lambda_1[A,B_1]+\lambda_2[A,B_2]$.

(3) $[A,B]=-[B,A]$.

(4) ***(Jacobi identity)*** $[[A,B],C]+[[B,C],A]+[[C,A],B]=0$.

The group operation in $G$ determines the Lie bracket operation in $\mathfrak{g}$. One therefore expects smoothly isomorphic groups to have isomorphic Lie algebras. Before proving this, we need to precisely define "isomorphic Lie algebras".

**Definition 8.5.** *Let* $\mathfrak{g}_1$ *and* $\mathfrak{g}_2$ *be the Lie algebras of two matrix groups. A linear transformation* $f:\mathfrak{g}_1\to\mathfrak{g}_2$ *is called a* ***Lie algebra homomorphism*** *if for all* $A,B\in\mathfrak{g}_1$*,*

$$f([A,B])=[f(A),f(B)].$$

*If* $f$ *is also bijective, then* $f$ *is called a* ***Lie algebra isomorphism****.*

The most important Lie algebra homomorphisms are the ones determined by smooth group homomorphisms.

**Proposition 8.6.** *Let* $G_1,G_2$ *be matrix groups with Lie algebras* $\mathfrak{g}_1,\mathfrak{g}_2$*. Let* $f:G_1\to G_2$ *be a smooth homomorphism. Then the derivative* $df_I:\mathfrak{g}_1\to\mathfrak{g}_2$ *is a Lie algebra homomorphism.*

**Proof.** Let $A, B \in \mathfrak{g}_1$. Let $t \mapsto a(t)$ and $t \mapsto b(t)$ be differentiable paths in $G_1$ with $a(0) = b(0) = I$, $a'(0) = A$ and $b'(0) = B$. Fix $t_0 \in \mathbb{R}$, and set $a = a(t_0) \in G_1$. We first show that:

$$df_I(\mathrm{Ad}_a(B)) = \mathrm{Ad}_{f(a)}(df_I(B)). \tag{8.2}$$

Since $\alpha(t) = ab(t)a^{-1}$ satisfies $\alpha(0) = I$ and $\alpha'(0) = \mathrm{Ad}_a(B)$, Equation 8.2 can be justified as follows:

$$\begin{aligned} df_I(\mathrm{Ad}_a(B)) &= \frac{d}{dt}\Big|_{t=0}(f \circ \alpha)(t) = \frac{d}{dt}\Big|_{t=0} f(ab(t)a^{-1}) \\ &= \frac{d}{dt}\Big|_{t=0} f(a)f(b(t))f(a)^{-1} = \mathrm{Ad}_{f(a)}(df_I(B)). \end{aligned}$$

This confirms Equation 8.2, which we now use as follows:

$$\begin{aligned} df_I([A,B]) &= df_I\left(\frac{d}{dt}\Big|_{t=0}\mathrm{Ad}_{a(t)}B\right) = \frac{d}{dt}\Big|_{t=0} df_I(\mathrm{Ad}_{a(t)}B) \\ &= \frac{d}{dt}\Big|_{t=0}\mathrm{Ad}_{f(a(t))}(df_I(B)) = [df_I(A), df_I(B)]. \end{aligned}$$

The second equality above implicitly uses that, since $df_I : \mathfrak{g}_1 \to \mathfrak{g}_2$ is linear, its derivative at any point of $\mathfrak{g}_1$ is itself (see Exercise 7.6). So, since $v(t) = Ad_{a(t)}B$ is a path in $\mathfrak{g}_1$, this second equality is justified by:

$$df_I(v'(0)) = d(df_I)_{v(0)}(v'(0)) = \frac{d}{dt}\Big|_{t=0} df_I(v(t)).$$

□

**Corollary 8.7.** *Smoothly isomorphic matrix groups have isomorphic Lie algebras.*

**Proof.** If $f : G_1 \to G_2$ is a smooth isomorphism between two matrix groups, then $df_I : \mathfrak{g}_1 \to \mathfrak{g}_2$ is a linear isomorphism by Exercise 7.10 and is a Lie algebra homomorphism by Proposition 8.6. □

The following familiar 3-dimensional matrix groups have mutually isomorphic Lie algebras:

(8.3)

$$so(3) = \text{span}\left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\}.$$

$$su(2) = \text{span}\left\{ \frac{1}{2}\begin{pmatrix} \mathbf{i} & 0 \\ 0 & -\mathbf{i} \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 0 & \mathbf{i} \\ \mathbf{i} & 0 \end{pmatrix} \right\}.$$

$$sp(1) = \text{span}\left\{ \frac{1}{2}(\mathbf{i}), \frac{1}{2}(\mathbf{j}), \frac{1}{2}(\mathbf{k}) \right\}.$$

In all three, for the given basis $\{A_1, A_2, A_3\}$, it is straightforward to check:

$$[A_1, A_2] = A_3, \quad [A_2, A_3] = A_1, \quad [A_3, A_1] = A_2.$$

So all three Lie algebras have the same Lie bracket structure, or more precisely, the linear map between any two of them that sends basis elements to corresponding basis elements is a Lie algebra isomorphism. If these bases are used to identify the Lie algebras with $\mathbb{R}^3$, notice that the Lie bracket operation becomes the familiar cross product from vector calculus.

The fact that $su(2) \cong sp(1)$ is not surprising, since $SU(2)$ and $Sp(1)$ are smoothly isomorphic (by Proposition 3.13). We will later learn that $SO(3)$ is neither isomorphic nor homeomorphic to $Sp(1)$, in spite of the fact that their Lie algebras look identical. Another such example is the pair $SO(n)$ and $O(n)$, which have identical Lie algebras but are not isomorphic. It turns out that path-connected, *simply connected* matrix groups are smoothly isomorphic if and only if their Lie algebras are isomorphic. It is beyond the scope of this text to precisely define "simply connected" or prove this fact.

## 2. The adjoint representation

Let $G \subset GL_n(\mathbb{K})$ be a matrix group of dimension $d$, with Lie algebra $\mathfrak{g}$. For every $g \in G$, $\text{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ is a vector space isomorphism. Once we choose a basis, $\mathcal{B}$, of $\mathfrak{g}$, this isomorphism can be represented as $L_A$ for some $A \in GL_d(\mathbb{R})$, as in Section 7 of Chapter 1. In other words, after fixing a basis of $\mathfrak{g}$, we can regard the map $g \mapsto \mathrm{Ad}_g$ as a function from $G$ to $GL_d(\mathbb{R})$.

**Lemma 8.8.** *$Ad: G \to GL_d(\mathbb{R})$ is a smooth homomorphism.*

**Proof.** For all $g_1, g_2 \in G$ and all $X \in \mathfrak{g}$,

$$\mathrm{Ad}_{g_1 g_2}(X) = (g_1 g_2)X(g_1 g_2)^{-1} = g_1 g_2 X g_2^{-1} g_1^{-1} = \mathrm{Ad}_{g_1}(\mathrm{Ad}_{g_2}(X)).$$

This shows that $\mathrm{Ad}_{g_1 g_2} = \mathrm{Ad}_{g_1} \circ \mathrm{Ad}_{g_2}$. Since the composition of two linear maps corresponds to the product of the matrices representing them, this verifies that $\mathrm{Ad} : G \to GL_d(\mathbb{R})$ is a homomorphism. We leave to the reader (in Exercise 8.11) the straightforward verification that Ad is smooth. $\square$

This homomorphism is called the **adjoint representation** of $G$ on $\mathfrak{g}$. As we will explore in Chapter 10, in general a **representation** of a matrix group $G$ on a Euclidean space $\mathbb{R}^m$ means a homomorphism from $G$ to $GL_m(\mathbb{R})$. It associates each element of $G$ with a linear transformation of $\mathbb{R}^m$, and hence determines how elements of $G$ "act on" vectors in $\mathbb{R}^m$. For example, we have studied all along how $SO(n)$ acts on $\mathbb{R}^n$; it is interesting that $SO(n)$ also acts naturally on $so(n) \cong \mathbb{R}^{n(n-1)/2}$.

The image of Ad in $GL_d(\mathbb{R})$ contains only Lie algebra isomorphisms, since:

**Lemma 8.9.** *For all $g \in G$ and all $X, Y \in \mathfrak{g}$,*

$$[Ad_g(X), Ad_g(Y)] = Ad_g([X, Y]).$$

**Proof.** This follows from Proposition 8.6, since $\mathrm{Ad}_g = d(C_g)_I$. An alternative proof is the following explicit verification:

$$\begin{aligned}
[\mathrm{Ad}_g(X), \mathrm{Ad}_g(Y)] &= [gXg^{-1}, gYg^{-1}] \\
&= gXg^{-1}gYg^{-1} - gYg^{-1}gXg^{-1} \\
&= g(XY - YX)g^{-1} \\
&= \mathrm{Ad}_g([X, Y]).
\end{aligned}$$

$\square$

The fact that $\text{Ad} : G \to GL_d(\mathbb{R})$ is a smooth homomorphism has a very strong consequence; namely, Ad sends one-parameter groups in $G$ to one-parameter groups in $GL_d(\mathbb{R})$. To elaborate on this comment, for any $X \in \mathfrak{g}$, we denote by

$$\text{ad}_X : \mathfrak{g} \to \mathfrak{g}$$

the linear map that sends $Y$ to $[X, Y]$. That is, $\text{ad}_X(Y) = [X, Y]$.

**Proposition 8.10.** *For all $X \in \mathfrak{g}$, $Ad_{e^X} = e^{ad_X}$.*

Before proving this proposition, we explain it. On the right side, exponentiation of the linear map $\text{ad}_X : \mathfrak{g} \to \mathfrak{g}$ is defined as follows. In our fixed basis, $\mathcal{B}$, of $\mathfrak{g}$, $\text{ad}_X$ is represented by a matrix. This matrix can be exponentiated, and the linear transformation $\mathfrak{g} \to \mathfrak{g}$ associated to the result is denoted $e^{\text{ad}_X}$. The result is independent of the choice of $\mathcal{B}$ by Proposition 6.18. In fact, $e^{\text{ad}_X}$ can be computed by formally substituting $\text{ad}_X$ into the exponential power series. That is, for all $Y \in \mathfrak{g}$,

$$\begin{aligned}(e^{\text{ad}_X})(Y) &= (I + (\text{ad}_X) + (1/2)(\text{ad}_X)^2 + (1/6)(\text{ad}_X)^3 + \cdots)Y \\ &= Y + [X, Y] + (1/2)[X, [X, Y]] + (1/6)[X, [X, [X, Y]]] + \cdots .\end{aligned}$$

So Proposition 8.10 says that the transformation $\text{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ (when $g = e^X$) can be calculated purely in terms of repeated Lie brackets with $X$.

**Proof.** The key is that for $X \in \mathfrak{g}$, $d(\text{Ad})_I(X) \in gl_d(\mathbb{R})$ is the matrix representing $\text{ad}_X$. We abbreviate this as:

$$d(\text{Ad})_I(X) = \text{ad}_X. \tag{8.4}$$

Equation 8.4 follows immediately from Definition 8.1, or more explicitly by observing that for all $Y \in \mathfrak{g}$:

$$\frac{d}{dt}\bigg|_{t=0} \text{Ad}_{e^{tX}}(Y) = \frac{d}{dt}\bigg|_{t=0} (e^{tX} Y e^{-tX}) = XY - YX = \text{ad}_X(Y). \tag{8.5}$$

Now, $\text{Ad} : G \to GL_d(\mathbb{R})$ is a smooth homomorphism. By Exercise 7.16, a smooth homomorphism between matrix groups sends one-parameter groups to one-parameter groups and is therefore completely determined by its derivative at $I$. More precisely, for all $X \in \mathfrak{g}$,

$$\mathrm{Ad}_{e^X} = e^{d(\mathrm{Ad})_I(X)} = e^{\mathrm{ad}_X}.$$

□

## 3. Example: the adjoint representation for $SO(3)$

In this section, we explicitly compute the adjoint representation for $SO(3)$. For this purpose, a convenient choice of basis of $so(3)$ is:

$$\left\{ E_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, E_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, E_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\}.$$

As mentioned in Equation 8.3, the Lie bracket structure is:

$$[E_1, E_2] = E_3, \;\; [E_2, E_3] = E_1, \;\; [E_3, E_1] = E_2.$$

This basis determines a vector space isomorphism $f : \mathbb{R}^3 \to so(3)$, namely,

$$(a, b, c) \overset{f}{\mapsto} \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}.$$

For every $g \in SO(3)$, $\mathrm{Ad}_g : so(3) \to so(3)$ can be regarded (via $f$) as a linear map $\mathbb{R}^3 \to \mathbb{R}^3$, which equals left-multiplication by some matrix. We carefully chose the basis above such that this matrix will turn out to be $g$. In other words, we will prove that conjugating an element of $so(3)$ by $g$ gives the same answer as left-multiplying the corresponding vector in $\mathbb{R}^3$ by $g$:

$$g(f(a, b, c))g^{-1} = f(L_g(a, b, c)). \tag{8.6}$$

Equation 8.6 is equivalent to the following proposition:

**Proposition 8.11.** *In the above basis, $Ad : SO(3) \to GL_3(\mathbb{R})$ is just the inclusion map, which sends every matrix to itself.*

**Proof.** We first show that the derivative $d(\mathrm{Ad})_I : so(3) \to gl_3(\mathbb{R})$ sends every matrix to itself. Let $\gamma(t)$ be a path in $SO(3)$ with $\gamma(0) = I$

and $\gamma'(0) = E_1$. Let $v \in so(3)$. Then $t \mapsto \mathrm{Ad}_{\gamma(t)}(v)$ is a path in $so(3)$ whose derivative equals (as in Equation 8.5):

$$\frac{d}{dt}\Big|_{t=0} \mathrm{Ad}_{\gamma(t)}(v) = \frac{d}{dt}\Big|_{t=0} \gamma(t)v\gamma(t)^{-1}$$

$$= E_1 v - vE_1 = [E_1, v] = \begin{cases} 0 \text{ if } v = E_1 \\ E_3 \text{ if } v = E_2 \\ -E_2 \text{ if } v = E_3 \end{cases} = f(L_{E_1}(f^{-1}(v))).$$

Thus, the linear transformation

$$v \mapsto \frac{d}{dt}\Big|_{t=0} \mathrm{Ad}_{\gamma(t)}(v)$$

is represented in this basis as left multiplication by the matrix $E_1$. This shows that $d(\mathrm{Ad})_I(E_1) = E_1$. A similar argument gives that $d(\mathrm{Ad})_I(E_2) = E_2$ and $d(\mathrm{Ad})_I(E_3) = E_3$. Thus, $d(\mathrm{Ad})_I$ sends every matrix in $so(3)$ to itself.

Since $d(\mathrm{Ad})_I$ sends every matrix to itself, $\mathrm{Ad} : SO(3) \to GL_3(\mathbb{R})$ sends every one-parameter group in $SO(3)$ to itself. To conclude that Ad sends all of $SO(3)$ to itself, one can use the following result from the next chapter: $\exp : so(3) \to SO(3)$ is surjective, so every element of $SO(3)$ is contained in a one-parameter group. Or to conclude the proof without using this fact, one can verify that the set of all $g \in SO(3)$ sent to themselves by Ad is clopen, and hence is all of $SO(3)$ (since $SO(3)$ is path-connected by Exercise 4.14). $\square$

## 4. The adjoint representation for compact matrix groups

We saw that the image of Ad in $GL_d(\mathbb{R})$ contains only Lie algebra isomorphisms. A second important restriction on the image of $Ad$ in $GL_d(\mathbb{R})$ applies only when $G$ is a subgroup of $O(n)$, $U(n)$ or $Sp(n)$:

**Proposition 8.12.** *If $G$ is a subgroup of $\mathcal{O}_n(\mathbb{K})$, then for all $g \in G$ and all $X \in \mathfrak{g}$, $|Ad_g(X)| = |X|$.*

Remember that $\mathfrak{g} \subset M_n(\mathbb{K}) \cong \mathbb{R}^{n^2}$, $\mathbb{R}^{2n^2}$ or $\mathbb{R}^{4n^2}$, and that $|\cdot|$ denotes the Euclidean norm on $M_n(\mathbb{K})$. Further, $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ denotes the Euclidean inner product on $M_n(\mathbb{K})$, which (as we saw in Section 1 of

Chapter 3) equals the real part of the $\mathbb{K}$-inner product, $\langle\cdot,\cdot\rangle_{\mathbb{K}}$. For example, in $u(2)$,

$$\left|\begin{pmatrix} a\mathbf{i} & b+c\mathbf{i} \\ -b+c\mathbf{i} & d\mathbf{i} \end{pmatrix}\right| = \sqrt{a^2+2b^2+2c^2+d^2},$$

and

$$(8.7)\quad \left\langle \begin{pmatrix} a_1\mathbf{i} & b_1+c_1\mathbf{i} \\ -b_1+c_1\mathbf{i} & d_1\mathbf{i} \end{pmatrix}, \begin{pmatrix} a_2\mathbf{i} & b_2+c_2\mathbf{i} \\ -b_2+c_2\mathbf{i} & d_2\mathbf{i} \end{pmatrix} \right\rangle_{\mathbb{R}} = a_1a_2+2b_1b_2+2c_1c_2+d_1d_2.$$

**Proof.** For $X, Y \in M_n(\mathbb{K}) \cong \mathbb{K}^{n^2}$, a convenient alternative description of their $\mathbb{K}$-inner product, $\langle X, Y\rangle_{\mathbb{K}}$, is:

$$(8.8)\qquad \langle X, Y\rangle_{\mathbb{K}} = \text{trace}(X\cdot Y^*).$$

Equation 8.8 is justified as follows:

$$\begin{aligned} \text{trace}(X\cdot Y^*) = \sum_{i=1}^{n}(X\cdot Y^*)_{ii} &= \sum_{i=1}^{n}\sum_{j=1}^{n} X_{ij}(Y^*)_{ji} \\ &= \sum_{i,j=1}^{n} X_{ij}\overline{Y}_{ij} = \langle X, Y\rangle_{\mathbb{K}}. \end{aligned}$$

We will use this alternative description to prove that for all $g \in \mathcal{O}_n(\mathbb{K})$ and all $X \in M_n(\mathbb{K})$,

$$(8.9)\qquad |Xg| = |gX| = |X|.$$

To justify Equation 8.9, we use the fact that $g\cdot g^* = I$:

$$|Xg|^2 = \text{trace}((Xg)(Xg)^*) = \text{trace}(Xgg^*X^*) = \text{trace}(XX^*) = |X|^2.$$

For the other half:

$$\begin{aligned} |gX|^2 = |(gX)^*|^2 = \text{trace}((gX)^*(gX)) &= \text{trace}(X^*g^*gX) \\ &= \text{trace}(X^*X) = |X^*|^2 = |X|^2. \end{aligned}$$

The proposition follows immediately, since $|gXg^{-1}| = |gX| = |X|$. $\square$

Equation 8.8 provides an important description of the $\mathbb{R}$-inner product of vectors $X, Y \in \mathfrak{g}$:

$$\langle X, Y\rangle_{\mathbb{R}} = \text{Real}\left(\langle X, Y\rangle_{\mathbb{K}}\right) = \text{Real}(\text{trace}(X \cdot Y^*)). \tag{8.10}$$

Proposition 8.12 can be restated as follows:

**Corollary 8.13.** *Assume that $G$ is a subgroup of $\mathcal{O}_n(\mathbb{K})$. If the fixed basis, $\mathcal{B}$, of $\mathfrak{g}$ is orthonormal with respect to $\langle \cdot, \cdot\rangle_{\mathbb{R}}$, then $Ad_g \in O(d)$ for all $g \in G$. Thus, $Ad : G \to O(d)$ is a homomorphism from $G$ to the orthogonal group $O(d)$.*

**Proof.** For all $g \in G$, $\text{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ preserves norms and therefore also inner products. That is, $\langle \text{Ad}_g(X), \text{Ad}_g(Y)\rangle_{\mathbb{R}} = \langle X, Y\rangle_{\mathbb{R}}$ for all $X, Y \in \mathfrak{g}$. The result now follows from Exercise 3.14. ☐

An important consequence is that Lie brackets interact with the $\mathbb{R}$-inner product in the following way:

**Proposition 8.14.** *If $G$ is a subgroup of $\mathcal{O}_n(\mathbb{K})$, then for all vectors $X, Y, Z \in \mathfrak{g}$,*

$$\langle [X, Y], Z\rangle_{\mathbb{R}} = -\langle [X, Z], Y\rangle_{\mathbb{R}}.$$

**Proof.** Let $\alpha(t)$ be a path in $G$ with $\alpha(0) = I$ and $\alpha'(0) = X$. Since $\langle \cdot, \cdot\rangle_{\mathbb{R}}$ is Ad-invariant:

$$\begin{aligned} 0 &= \left.\frac{d}{dt}\right|_{t=0} \left\langle \text{Ad}_{\alpha(t)} Y, \text{Ad}_{\alpha(t)} Z\right\rangle_{\mathbb{R}} \\ &= \left\langle \left.\frac{d}{dt}\right|_{t=0} \text{Ad}_{\alpha(t)} Y, Z\right\rangle_{\mathbb{R}} + \left\langle Y, \left.\frac{d}{dt}\right|_{t=0} \text{Ad}_{\alpha(t)} Z\right\rangle_{\mathbb{R}} \\ &= \langle [X, Y], Z\rangle_{\mathbb{R}} + \langle Y, [X, Z]\rangle_{\mathbb{R}}. \end{aligned}$$

The second equality uses the rule $\langle A, B\rangle' = \langle A', B\rangle + \langle A, B'\rangle$, which is a basic differentiation rule for the dot product found in any multi-variable calculus textbook. ☐

We end this section by looking more carefully at Equation 8.9, which said that for all $g \in \mathcal{O}_n(\mathbb{K})$ and all $X \in M_n(\mathbb{K})$,

$$|Xg| = |gX| = |X|.$$

We learned back in Chapter 3 that left or right multiplication by $g$ determined an isometry of $\mathbb{K}^n \cong \mathbb{R}^n, \mathbb{R}^{2n}$ or $\mathbb{R}^{4n}$. Now we learn that left or right multiplication by $g$ also determines an isometry of $M_n(\mathbb{K}) \cong \mathbb{R}^{n^2}, \mathbb{R}^{2n^2}$ or $\mathbb{R}^{4n^2}$.

One consequence is that a subgroup $G \subset \mathcal{O}_n(\mathbb{K}) \subset M_n(\mathbb{K})$ tends to have many symmetries. In fact, for any $g \in G$, the function $x \mapsto gx$ (or $x \mapsto xg$) is an isometry of $M_n(\mathbb{K})$ that carries $G$ to itself; in other words, it is a *symmetry* of $G$, as defined in Section 7 of Chapter 3. Thus, subgroups of $\mathcal{O}_n(\mathbb{K})$ tend to be highly symmetric manifolds.

## 5. Global conclusions

By definition, the Lie bracket provides information about the group operation among elements near $I$. What about elements far from $I$? In this section, we demonstrate some global conclusion about a group that can be derived purely from information about its Lie algebra.

Let $G$ be a matrix group with Lie algebra $\mathfrak{g}$. A subspace $\mathfrak{h} \subset \mathfrak{g}$ is called a **subalgebra** if it is closed under the Lie bracket operation; that is, $[A, B] \in \mathfrak{h}$ for all $A, B \in \mathfrak{h}$. Further, $\mathfrak{h}$ is called an **ideal** if $[A, B] \in \mathfrak{h}$ for all $A \in \mathfrak{h}$ and $B \in \mathfrak{g}$. Notice that the Lie algebra of any subgroup of $G$ is a subalgebra of $\mathfrak{g}$. We will prove:

**Theorem 8.15.** *Let $G$ be a path-connected matrix group, and let $H \subset G$ be a closed path-connected subgroup. Denote their Lie algebras as $\mathfrak{h} \subset \mathfrak{g}$. Then $H$ is a normal subgroup of $G$ if and only if $\mathfrak{h}$ is an ideal of $\mathfrak{g}$.*

**Proof.** First assume that $H$ is a normal subgroup of $G$. Let $A \in \mathfrak{h}$ and $B \in \mathfrak{g}$. Let $a(t)$ be a path in $H$ with $a(0) = I$ and $a'(0) = A$. Let $b(t)$ be a path in $G$ with $b(0) = I$ and $b'(0) = B$.

$$
\begin{aligned}
[A, B] = -[B, A] = -\frac{d}{dt}\Big|_{t=0} Ad_{b(t)} A& \\
&= -\frac{d}{dt}\Big|_{t=0} \left( \frac{d}{ds}\Big|_{s=0} b(t)a(s)b(t)^{-1} \right).
\end{aligned}
$$

Since $H$ is normal in $G$, $b(t)a(s)b(t)^{-1} \in H$, which implies $[A, B] \in \mathfrak{h}$.

Next assume that $\mathfrak{h}$ is an ideal of $\mathfrak{g}$. For every $B \in \mathfrak{g}$ and every $A \in \mathfrak{h}$,

$$\mathrm{Ad}_{e^B} A \in \mathfrak{h},$$

because it is the limit of a series of elements of $\mathfrak{h}$ by Theorem 8.10:

$$\begin{aligned}\mathrm{Ad}_{e^B} A &= e^{\mathrm{ad}_B}(A)\\ &= A + [B, A] + (1/2)[B, [B, A]] + (1/6)[B, [B, [B, A]]] + \cdots .\end{aligned}$$

Now let $a \in H$ and $b \in G$. Assume that $a$ and $b$ lie in a small neighborhood, $U$, of $I$ in $G$, so that $a = e^A$ for some $A \in \mathfrak{h}$ and $b = e^B$ for some $B \in \mathfrak{g}$. Then

$$bab^{-1} = be^A b^{-1} = e^{bAb^{-1}} = e^{Ad_b(A)} \in H.$$

We leave the reader (in Exercise 8.3) to show that $bab^{-1} \in H$ for all $a \in H$ and $b \in G$ (not necessarily close to $I$). □

The previous proof demonstrates that it is possible to derive a global conclusion about a matrix group ($H$ is normal in $G$) from a hypothesis about its Lie algebra ($\mathfrak{h}$ is an ideal of $\mathfrak{g}$). The Lie algebra, with its Lie bracket operation, seems to encode a lot of information about the matrix group. It turns out that the Lie bracket operation in $\mathfrak{g}$ completely determines the group operation in $G$, at least in a neighborhood of the identity! An explicit verification of this surprising claim is provided by the **Campbell-Baker-Hausdorff series**. For $X, Y, Z \in \mathfrak{g}$ with sufficiently small norm, the equation $e^X e^Y = e^Z$ has a power series solution for $Z$ in terms of repeated Lie brackets of $X$ and $Y$. The beginning of the series is:

$$Z = X + Y + (1/2)[X, Y] + (1/12)[X, [X, Y]] + (1/12)[Y, [Y, X]] + \cdots .$$

The existence of such a series means that the group operation is completely determined by the Lie bracket operation; the product of $e^X$ and $e^Y$ equals $e^Z$, where $Z$ can be expressed purely in terms of repeated Lie brackets of $X$ and $Y$.

We conclude this section with a caution: the Lie algebra, $\mathfrak{g}$, of a matrix group, $G$, contains information only about the identity component, $G_0$, of $G$ (defined in Exercise 7.7). For example, $G = SL_n(\mathbb{Z})$ (defined in Exercise 1.8) has identity component $G_0 = \{I\}$ and Lie algebra $\mathfrak{g} = \{0\}$. This matrix group is comprised of discrete points; the Lie algebra tells you nothing about the interesting group operation on this set of discrete points.

## 6. The double cover $Sp(1) \to SO(3)$

In this section, we study the adjoint representation of $Sp(1)$:

$$\text{Ad} : Sp(1) \to O(3).$$

Since $Sp(1)$ is path-connected (by Exercise 4.15), so is its image under Ad (by Exercise 4.16), so we in fact have a smooth homomorphism:

$$\text{Ad} : Sp(1) \to SO(3).$$

Our goal is to prove that $\text{Ad} : Sp(1) \to SO(3)$ is a surjective, 2-to-1 local diffeomorphism. The term **local diffeomorphism** means that there exists a neighborhood of any point of the domain, restricted to which the function is a diffeomorphism onto its image. A surjective 2-to-1 local diffeomorphism between compact manifolds is often called a **double cover**. This double cover provides an extremely useful tool for better understanding both $Sp(1)$ and $SO(3)$.

For $g \in Sp(1)$ (regarded as a unit-length quaternion) and for $v \in sp(1) = \text{span}\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, we have that

$$Ad_g(v) = gvg^{-1} \in sp(1).$$

Notice that conjugation by $g$ determines an isometry $\mathbb{H} \to \mathbb{H}$, which fixes $\text{span}\{1\}$ and thus also fixes $sp(1) = \text{span}\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$. So the adjoint representation of $Sp(1)$ can be regarded as conjugation restricted to the purely imaginary quaternions. Therefore $Ad$ equals the homomorphism that we named $\varphi$ in Exercise 3.16 and continued to study in Exercises 4.27 and 7.18. In fact, the following three lemmas are nothing but solutions to those exercises.

**Lemma 8.16.** *$Ker(Ad) = \{1, -1\}$, and therefore Ad is 2-to-1.*

**Proof.** If $g \in \text{Ker(Ad)}$, then $gvg^{-1} = v$ for all $v \in sp(1)$. In other words, $g$ commutes with all purely imaginary quaternions, and hence with all quaternions. So $g \in \mathbb{R}$ by Exercise 1.18, which means that $g = \pm 1$. □

**Lemma 8.17.** *Ad is a local diffeomorphism at $I$. In other words, Ad restricted to a sufficiently small neighborhood of $I$ in $Sp(1)$ is a diffeomorphism onto its image.*

**Proof.** By the inverse function theorem for manifolds (7.22), it will suffice to prove that $d(\mathrm{Ad})_I : sp(1) \to so(3)$ sends the natural basis $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ of $sp(1)$ to a basis of $so(3)$.

The path $\gamma(t) = e^{\mathbf{i}t} = \cos(t) + \mathbf{i}\sin(t)$ in $Sp(1)$ satisfies $\gamma(0) = I$ and $\gamma'(0) = \mathbf{i}$. For all $v \in sp(1) = \mathrm{span}\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$,

$$\frac{d}{dt}\bigg|_{t=0} \mathrm{Ad}_{\gamma(t)}(v) = \frac{d}{dt}\bigg|_{t=0} e^{it} v e^{-it} = \mathbf{i}v - v\mathbf{i} = \begin{cases} 0 \text{ if } v = \mathbf{i} \\ 2\mathbf{k} \text{ if } v = \mathbf{j} \\ -2\mathbf{j} \text{ if } v = \mathbf{k}. \end{cases}$$

This shows that

$$d(\mathrm{Ad})_I(\mathbf{i}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -2 \\ 0 & 2 & 0 \end{pmatrix}.$$

Now repeat this argument with $\mathbf{j}$ and $\mathbf{k}$ to verify that

$$\{d(\mathrm{Ad})_I(\mathbf{i}), d(\mathrm{Ad})_I(\mathbf{j}), d(\mathrm{Ad})_I(\mathbf{k})\}$$

is a basis of $so(3)$. □

It follows from Exercise 8.2 at the end of this chapter that Ad is a local diffeomorphism at every $g \in Sp(1)$ (not just at the identity).

We prove next that every element of $SO(3)$ is in the image of Ad. This might be surprising, since elements in the image are all Lie algebra isomorphisms of $sp(1)$. But this restriction is redundant, since matrices in $SO(3)$ preserve the vector cross-product in $\mathbb{R}^3$, which is the same as the Lie bracket operation in $sp(1)$, and are therefore automatically Lie algebra isomorphisms.

**Lemma 8.18.** *$Ad\colon Sp(1) \to SO(3)$ is surjective.*

**Proof.** Since $Sp(1)$ is compact (Proposition 4.28), its image under Ad is compact (by Proposition 4.24) and therefore closed. On the other hand, this image is open by the local diffeomorphism property. Thus, the image is a non-empty clopen subset of $SO(3)$. Since $SO(3)$ is path-connected (by Exercise 4.14), its only non-empty clopen subset is all of $SO(3)$ (see Proposition 4.17). □

This double cover $Sp(1) \to SO(3)$ has many implications. It explains why $Sp(1)$ and $SO(3)$ have isomorphic Lie algebras. Its algebraic import can be summarized as follows:

$$SO(3) \text{ is isomorphic to } Sp(1)/\{1, -1\},$$

which makes sense because $\{1, -1\}$ is a normal subgroup of $Sp(1)$.

Its geometric import has to do with the shape of $SO(3)$. We will show that $SO(3)$ is diffeomorphic to an important manifold called $\mathbb{RP}^3$.

**Definition 8.19.** *The set of all lines through the origin in $\mathbb{R}^{n+1}$ is called $n$-dimensional* ***real projective space*** *and is denoted as $\mathbb{RP}^n$.*

Since every line through the origin in $\mathbb{R}^{n+1}$ intersects the sphere $S^n$ in a pair of antipodal points, one often identifies $\mathbb{RP}^n$ with the set of antipodal pairs on $S^n$. The isomorphism $SO(3) \cong Sp(1)/\{I, -I\}$ associates each point of $SO(3)$ with a pair of antipodal points on the sphere $S^3 \cong Sp(1)$, and therefore provides a bijection between $SO(3)$ and $\mathbb{RP}^3$. This natural bijection helps us understand the shape of $SO(3)$.

It may seem inappropriate that we referred to $\mathbb{RP}^n$ as a manifold, since it is not a subset of any Euclidean space. Chapter 10 contains a more general definition of "manifold" under which $\mathbb{RP}^n$ can be proven to be one. For now, it suffices to just regard $\mathbb{RP}^3$ as a manifold by identifying it with $SO(3)$, as above.

We learned in Exercise 7.14 that $SO(3)$ is diffeomorphic to $T^1S^2$, which is another way to visualize the shape of $SO(3)$. Using some topology that is beyond the scope of this book, one can prove:

**Proposition 8.20.** *No pair of the following three 3-dimensional manifolds is homeomorphic:*

(1) $T^1S^2 = SO(3) = \mathbb{RP}^3$,

(2) $S^3 = Sp(1) = SU(2)$,

(3) $S^2 \times S^1$.

In particular, $SO(3)$ is not homeomorphic to $S^2 \times S^1$, which implies a negative answer to Question 1.2 from Chapter 1. The fact that

$T^1S^2$ is different from $S^2 \times S^1$ has practical implications for navigation on the earth. Because of this difference, it is impossible to construct a continuously changing basis for all of the tangent spaces of $S^2$. For example, the "east and north" basis does not extend continuously over the north and south poles.

The double cover $\mathrm{Ad} : Sp(1) \to SO(3)$ can be used to construct important finite groups if $Sp(1)$. If $H \subset SO(3)$ is a finite subgroup (these are classified in Section 7 of Chapter 3), then

$$\mathrm{Ad}^{-1}(H) = \{q \in Sp(1) \mid \mathrm{Ad}_q \in H\}$$

is a finite subgroup of $Sp(1)$ with twice the order of $H$. For example let $H \subset SO(3)$ denote the proper symmetry group of the icosahedron, which is isomorphic to $A_5$ (see Section 7 of Chapter 3). Let $H^* = \mathrm{Ad}^{-1}(H) \subset Sp(1)$. $H$ has order 60 and is called the *icosahedral group*. $H^*$ has order 120 and is called the *binary icosahedral group*. The set of cosets $Sp(1)/H^*$ is a three-dimensional manifold called the *Poincaré dodecahedral space*. This manifold is recently of interest to cosmologists because it has been proposed as a good candidate for the shape of the universe [**14**].

## 7. Other double covers

It turns out that for every $n > 2$, there is a matrix group that double-covers $SO(n)$. The first few are:

$$\begin{aligned} Sp(1) &\to SO(3) \\ Sp(1) \times Sp(1) &\to SO(4) \\ Sp(2) &\to SO(5) \\ SU(4) &\to SO(6) \end{aligned}$$

In general, the double cover of $SO(n)$ is denoted $\mathrm{Spin}(n)$ and is called the **spin group**, not to be confused with the symplectic group, $Sp(n)$. For $3 \le n \le 6$, $\mathrm{Spin}(n)$ is as above. For $n > 6$, $\mathrm{Spin}(n)$ is not isomorphic to any thus far familiar matrix groups. See [**3**] for a construction of the spin groups.

Since these double covers are group homomorphisms, the Lie algebra of Spin$(n)$ is isomorphic to the Lie algebra of $SO(n)$. Thus,

$$\begin{aligned} sp(1) &\cong so(3) \\ sp(1) \times sp(1) &\cong so(4) \\ sp(2) &\cong so(5) \\ su(4) &\cong so(6) \end{aligned}$$

We will describe only the second double cover above, denoted

$$F : Sp(1) \times Sp(1) \to SO(4).$$

Remember that $Sp(1) \times Sp(1)$ is a matrix group by Exercise 1.10. The double cover is defined such that for $(g_1, g_2) \in Sp(1) \times Sp(1)$ and $v \in \mathbb{R}^4 \cong \mathbb{H}$,

$$F(g_1, g_2)(v) = g_1 v \overline{g}_2.$$

This is exactly the homomorphism we defined in Exercise 3.17 and continued to study in Exercises 4.28 and 7.19. The following arguments are nothing but solutions to those exercises.

By arguments completely analogous to the previous section, the image of $F$ is $SO(4)$, and $F$ is a smooth 2-to-1 homomorphism and a local diffeomorphism at every point. Its kernel is $\{(I, I), (-I, -I)\}$, which means that:

$$SO(4) \cong (Sp(1) \times Sp(1))/\{(I, I), (-I, -I)\}.$$

The derivative $dF_{(I,I)} : sp(1) \times sp(1) \to so(4)$ is the following Lie algebra isomorphism:

$$dF_{(I,I)}(a\mathbf{i}+b\mathbf{j}+c\mathbf{k}, x\mathbf{i}+y\mathbf{j}+z\mathbf{k}) = \begin{pmatrix} 0 & -a-x & -b-y & -c-z \\ a+x & 0 & -c+z & b-y \\ b+y & c-z & 0 & -a+x \\ c+z & -b+y & a-x & 0 \end{pmatrix}.$$

This is straightforward to verify on basis elements. For example,

$$\frac{d}{dt}\Big|_{t=0} F(e^{\mathbf{i}t}, 1)(v) = \frac{d}{dt}\Big|_{t=0} e^{it} v = \mathbf{i}v = \begin{cases} \mathbf{i} \text{ if } v = 1 \\ -1 \text{ if } v = \mathbf{i} \\ \mathbf{k} \text{ if } v = \mathbf{j} \\ -\mathbf{j} \text{ if } v = \mathbf{k} \end{cases}$$

which shows that

$$dF_{(I,I)}(\mathbf{i},0) = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The vectors $(\mathbf{j},0), (\mathbf{k},0), (0,\mathbf{i}), (0,\mathbf{j})$ and $(0,\mathbf{k})$ are handled similarly.

The fact that $so(4)$ is isomorphic to $sp(1) \times sp(1)$ has many important consequences. It is the essential starting point on which the inter-related theories of 4-dimensional manifolds, Yang-Mills connections, and particle physics are built.

## 8. Exercises

**Ex. 8.1.** Question 1.1 in Chapter 1 asked whether $SO(3)$ is an abelian group. Prove that it is not in two ways: first by finding two elements of $so(3)$ that do not commute, and second by finding two elements of $SO(3)$ that do not commute. Which is easier? Prove that $SO(n)$ is not abelian for any $n > 2$.

**Ex. 8.2.** Let $G_1, G_2$ be matrix groups with Lie algebras $\mathfrak{g}_1, \mathfrak{g}_2$. Suppose that $f : G_1 \to G_2$ is a smooth homomorphism. If $df_I : \mathfrak{g}_1 \to \mathfrak{g}_2$ is bijective, prove that $df_g : T_g G_1 \to T_{f(g)} G_2$ is bijective for all $g \in G_1$.

**Ex. 8.3.**

(1) Let $G$ be a path-connected matrix group, and let $U$ be a neighborhood of $I$ in $G$. Prove that $U$ generates $G$, which means that every element of $G$ is equal to a finite product $g_1 g_2 \cdots g_k$, where for each $i$, $g_i$ or $g_i^{-1}$ lies in $U$.

(2) In the proof of Theorem 8.15, remove the restriction that $a$ and $b$ are close to the identity.

**Ex. 8.4.** Define $d : Sp(1) \to Sp(1) \times Sp(1)$ as $d(a) = (a,a)$. Explicitly describe the function $\iota : SO(3) \to SO(4)$ for which the following diagram commutes:

$$\begin{tikzcd} Sp(1) \arrow[r, "d"] \arrow[d, "\mathrm{Ad}"'] & Sp(1) \times Sp(1) \arrow[d, "F"] \\ SO(3) \arrow[r, "\iota"] & SO(4) \end{tikzcd}$$

**Ex. 8.5.** Express $so(4)$ as the direct sum of two 3-dimensional subspaces, each of which is an ideal of $so(4)$. Show there is a unique way to do so.

**Ex. 8.6.** Prove that $Sp(1) \times SO(3)$ is not smoothly isomorphic to $SO(4)$. *Hint: A smooth isomorphism would be determined by its derivative at $(I, I)$, which would send ideals to ideals.*

**Ex. 8.7.** Construct an explicit diffeomorphism between $SO(4)$ and $Sp(1) \times SO(3)$.

**Ex. 8.8.** Does there exist a basis for $u(2)$ such that the function $\text{Ad} : U(2) \to O(4)$ is the familiar injective map, denoted as $\rho_2$ in Chapter 2?

**Ex. 8.9.** Let $G$ be a path-connected matrix group, and let $H \subset G$ be a path-connected subgroup. Denote their Lie algebras as $\mathfrak{h} \subset \mathfrak{g}$. $H$ is called **central** if $gh = hg$ for all $g \in G$ and $h \in H$. Prove that $H$ is central if and only if $[X, Y] = 0$ for all $X \in \mathfrak{g}$ and $Y \in \mathfrak{h}$.

**Ex. 8.10.** Do $SO(3)$ and $\text{Isom}(\mathbb{R}^2)$ have isomorphic Lie algebras?

**Ex. 8.11.** For a matrix group $G$ of dimension $d$, prove that the function $\text{Ad} : G \to GL_d(\mathbb{R})$ is smooth.

**Ex. 8.12.** Let $G$ be a matrix group with Lie algebra $\mathfrak{g}$ and $A_1, A_2 \in \mathfrak{g}$.

(1) Prove that the path
$$\gamma(t) = e^{tA_1} e^{tA_2} e^{-tA_1} e^{-tA_2}$$
satisfies $\gamma(0) = I$, $\gamma'(0) = 0$ and $\gamma''(0) = 2[A_1, A_2]$. This is another precise sense in which $[A_1, A_2]$ measures the failure of $e^{tA_1}$ and $e^{tA_2}$ to commute for small $t$.

(2) Explicitly verify this when $\{A_1, A_2\}$ is a natural basis for the Lie algebra of $\text{Aff}_1(\mathbb{R})$. Explain visually in terms of translations and scalings of $\mathbb{R}$.

(3) Use a computer algebra system to explicitly verify this when $G = SO(3)$, and $A_1, A_2$ are the first two elements of the basis of $so(3)$ in Equation 8.3. In this example, is $\gamma(t)$ a one-parameter group? Explain this result in terms of rotations of a globe.

**Ex. 8.13.** Let $G$ be a closed subgroup of $O(n)$, $U(n)$ or $Sp(n)$. Let $\mathfrak{g}$ be the Lie algebra of $G$. Let $\mathfrak{h} \subset \mathfrak{g}$ be a subalgebra, and denote

$$\mathfrak{h}^\perp = \{A \in \mathfrak{g} \mid \langle X, A \rangle = 0 \text{ for all } X \in \mathfrak{h}\}.$$

(1) If $X \in \mathfrak{h}$ and $A \in \mathfrak{h}^\perp$, prove that $[X, A] \in \mathfrak{h}^\perp$.

(2) If $\mathfrak{h}$ is an ideal, prove that $\mathfrak{h}^\perp$ is also an ideal.

**Ex. 8.14.** In contrast to the fact that $T^1S^2 \neq S^2 \times S^1$, prove that $T^1S^3$ is diffeomorphic to $S^3 \times S^2$, and more generally that $T^1G$ is diffeomorphic to $G \times S^{d-1}$ for any matrix group $G$ of dimension $d$.

Chapter 9

# Maximal tori

In Chapter 1, we regarded $SO(3)$ as the group of positions of a globe. We asked whether every element of $SO(3)$ can be achieved, starting at the identity position, by rotating through some angle about some single axis. In other words, is every element of $SO(3)$ just a rotation? In this chapter, we provide an affirmative answer. Much more generally, we characterize elements of $SO(n)$, $SU(n)$, $U(n)$ and $Sp(n)$. An element of any of these groups is just a simultaneous rotation in a collection of orthogonal planes.

To explain and prove this characterization, we must understand *maximal tori*, a fundamental tool for studying compact matrix groups. We will use maximal tori in this chapter to prove several important theorems about familiar compact matrix groups, including:

**Theorem 9.1.** *Let $G \in \{SO(n), U(n), SU(n), Sp(n)\}$.*

(1) *Every element of $G$ equals $e^X$ for some $X$ in the Lie algebra of $G$.*

(2) *$G$ is path-connected.*

Notice that part (2) follows from part (1), since every element of $G$ is connected to the identity by a one-parameter group. It turns out that part (1) is true when $G$ is any *compact* path-connected matrix group, but is false for several *non-compact* path-connected matrix groups, like $SL_2(\mathbb{R})$ and $SL_2(\mathbb{C})$.

## 1. Several characterizations of a torus

In this section, we define a torus and prove that tori are the only path-connected compact abelian matrix groups.

Remember that $U(1) = \{(e^{\mathbf{i}\theta}) \mid \theta \in [0, 2\pi)\}$ is the circle-group whose group operation is addition of angles. $U(1)$ is abelian, path-connected, and isomorphic to $SO(2)$.

**Definition 9.2.** *The n-dimensional **torus** $T^n$ is the group*

$$T^n = \underbrace{U(1) \times U(1) \times \cdots \times U(1)}_{n \text{ copies}}.$$

In general, the product of two or more matrix groups is isomorphic to a matrix group by Exercise 1.10. In this case,

$$T^n \cong \{\text{diag}(e^{\mathbf{i}\theta_1}, ..., e^{\mathbf{i}\theta_n}) \mid \theta_i \in [0, 2\pi)\} \subset GL_n(\mathbb{C}).$$

There is a useful alternative description of $T^n$. Remember $(\mathbb{R}^n, +)$ denotes the group of vectors in Euclidean space under the operation of vector addition. $(\mathbb{R}^n, +)$ is isomorphic to a matrix group, namely $\text{Trans}(\mathbb{R}^n)$, as explained in Section 6 of Chapter 3.

In group theory, if $a_1, ..., a_k$ are elements of a group, $G$, one often denotes the subgroup of $G$ that they generate as $\langle\{a_1, ..., a_k\}\rangle \subset G$. This means the group of all finite products of the $a$'s and their inverses. For example, if $\{v_1, ..., v_k\} \subset (\mathbb{R}^n, +)$, then

$$\langle\{v_1, ..., v_k\}\rangle = \{n_1 v_1 + \cdots + n_k v_k \mid n_i \in \mathbb{Z}\},$$

where $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$ denotes the integers. Since $(\mathbb{R}^n, +)$ is an abelian group, any subgroup $N \subset (\mathbb{R}^n, +)$ is normal, so the coset space $(\mathbb{R}^n, +)/N$ is a group.

**Proposition 9.3.** *If $\{v_1, ..., v_n\} \subset \mathbb{R}^n$ is a basis, then the quotient group $(\mathbb{R}^n, +)/\langle\{v_1, ..., v_n\}\rangle$ is isomorphic to $T^n$.*

**Proof.** We first prove the proposition for the standard basis of $\mathbb{R}^n$,

$$\{e_1 = (1, 0, ..., 0), e_2 = (0, 1, 0, ..., 0), ..., e_n = (0, ..., 0, 1)\}.$$

The homomorphism $f : (\mathbb{R}^n, +) \to T^n$ defined as

$$f(t_1, ..., t_n) = \text{diag}(e^{2\pi\mathbf{i}t_1}, ..., e^{2\pi\mathbf{i}t_n})$$

is surjective. The kernel of $f$ equals $\langle\{e_1, ..., e_n\}\rangle$. Therefore, $T^n$ is isomorphic to $(\mathbb{R}^n, +)/\langle\{e_1, ..., e_n\}\rangle$.

Next let $\{v_1, ..., v_n\}$ be any basis of $\mathbb{R}^n$. Let $A \in GL_n(\mathbb{R})$ have rows equal to $v_1, ..., v_n$, so that $R_A(e_i) = v_i$ for all $i = 1, ..., n$. The function $R_A : (\mathbb{R}^n, +) \to (\mathbb{R}^n, +)$ is an isomorphism that sends the subgroup generated by the $e$'s to the subgroup generated by the $v$'s. It follows that

$$(\mathbb{R}^n, +)/\langle\{v_1, ..., v_n\}\rangle \cong (\mathbb{R}^n, +)/\langle\{e_1, ..., e_n\}\rangle \cong T^n.$$

□

**Corollary 9.4.** *If $\{v_1, ..., v_k\} \subset \mathbb{R}^n$ is a linearly independent set, then $(\mathbb{R}^n, +)/\langle\{v_1, ..., v_k\}\rangle$ is isomorphic to $T^k \times (\mathbb{R}^{n-k}, +)$.*

**Proof.** Choosing vectors $v_{k+1}, ..., v_n$ so that the $v$'s form a basis of $\mathbb{R}^n$,

$$\begin{aligned}
&(\mathbb{R}^n, +)/\langle\{v_1, ..., v_k\}\rangle \\
&\cong ((\text{span}\{v_1, ..., v_k\}, +)/\langle\{v_1, ..., v_k\}\rangle) \times (\text{span}\{v_{k+1}, ..., v_n\}, +) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \cong T^k \times (\mathbb{R}^{n-k}, +).
\end{aligned}$$

□

The term "torus" is justified by an important way to visualize $T^2$. In Figure 1, the collection of white circles is the subgroup $\langle\{v_1, v_2\}\rangle \subset \mathbb{R}^2$. In the quotient group, $\mathbb{R}^2/\langle\{v_1, v_2\}\rangle$, the coset containing a typical vector $w \in \mathbb{R}^2$ is pictured as a collection of grey circles. For most choices of $w \in \mathbb{R}^2$, this coset will intersect a **fundamental domain** (like the pictured parallelogram) in exactly one point. However, there are some exceptions. If $w \in \langle\{v_1, v_2\}\rangle$, then the coset intersects the fundamental domain in its four corners. If $w = nv_1 + tv_2$ for some $n \in \mathbb{Z}$ and some $t \in \mathbb{R}$ ($t \notin \mathbb{Z}$), then the coset intersects the fundamental domain in two points, one on its left and one on its right edge. Similarly, if $w = tv_1 + nv_2$, then the coset intersects the bottom and top edge of the fundamental domain.

Thus, $T^2$ can be identified with the fundamental domain, with the understanding that a point, $w$, on its left edge is considered the same as the point $w + v_1$ on its right edge, and similarly a point, $w$,

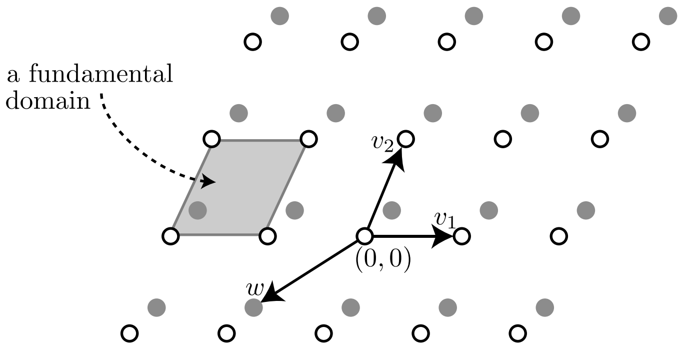



**Figure 1.** $T^2 = \mathbb{R}^2/\langle\{v_1, v_2\}\rangle$.

on its bottom edge is considered the same as the point $w + v_2$ on its top edge. If you cut a parallelogram out of stretchable material, and glue its left edge to its right edge, you obtain a cylinder. If you then glue the top edge to the bottom edge, you obtain something like the torus of revolution illustration in Section 4 of Chapter 7.

It is easy to see that $T^n$ is compact, abelian, and path-connected. We end this section by proving that these properties characterize tori.

**Theorem 9.5.** *Any compact, abelian path-connected matrix group, $G$, is isomorphic to a torus.*

**Proof.** Let $\mathfrak{g}$ denote the Lie algebra of $G$. Since $G$ is abelian, we have $AB = BA$, and therefore $e^A e^B = e^{A+B}$ for all $A, B \in \mathfrak{g}$. This equality means that the exponential map, $\exp : \mathfrak{g} \to G$, is a group homomorphism, when $\mathfrak{g}$ is considered as a group under vector addition.

Let $K \subset \mathfrak{g}$ denote the kernel of exp. $K$ is a **discrete subgroup** of $\mathfrak{g}$, which means that there exists a neighborhood, $U$, of the origin (= the identity) in $\mathfrak{g}$ whose intersection with $K$ contains only the origin, namely, any neighborhood on which exp is a diffeomorphism. It follows that any vector $v \in K$ has a neighborhood in $\mathfrak{g}$ separating it from all other elements of $K$; namely, $v + U = \{v + u \mid u \in U\}$.

Since $K$ is discrete, we claim that $K = \langle\{v_1, ..., v_k\}\rangle$ for some linearly independent set $\{v_1, ..., v_k\} \subset \mathfrak{g}$. For clarity, we will only indicate the argument when $\dim(\mathfrak{g}) = 2$, although the idea generalizes to any dimension.

Let $v_1 \in K$ denote a non-zero vector of minimal norm. Such a vector exists because $K$ is discrete. If $\langle\{v_1\}\rangle = K$, then the claim is true. Otherwise, let $v_2$ be a vector whose projection orthogonal to $v_1$ has minimal norm among candidates in $K$ but not in $\langle\{v_1\}\rangle$. Since $K$ is a subgroup, $\langle\{v_1, v_2\}\rangle \subset K$. The minimal-norm assumption for $v_1$ implies that $v_1$ and $v_2$ are linearly independent. We claim that $\langle\{v_1, v_2\}\rangle = K$. Suppose to the contrary that some vector $w \in K$ is not contained in $\langle\{v_1, v_2\}\rangle$. Then the four dashed-line-vectors pictured in Figure 2 are contained in $K$. It is straightforward to check that at least one of the four is too short, meaning it contradicts the minimal-norm property of $v_1$ or $v_2$.

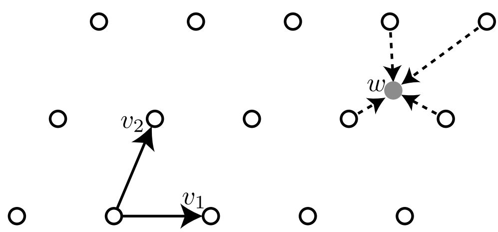



**Figure 2.** Proof that $\langle v_2, v_2\rangle = K$.

Thus, there exists a linearly independent set $\{v_1, ..., v_k\} \subset \mathfrak{g}$ such that $K = \langle\{v_1, .., v_k\}\rangle$.

Next we will prove that $\exp : \mathfrak{g} \to G$ is surjective. The image, $\exp(\mathfrak{g}) \subset G$, contains a neighborhood, $V$, of the identity in $G$. But this image is also a subgroup of $G$, since exp is a homomorphism. Therefore, $\exp(\mathfrak{g})$ contains the set, $\langle V\rangle$, consisting of all products of finitely many elements of $V$ and their inverses. We prove now that $\langle V\rangle = G$ (which amounts to solving part (1) of Exercise 8.3). Since $G$ is path-connected, it will suffice by Proposition 4.17 to prove that $\langle V\rangle$ is clopen in $G$. First, $\langle V\rangle$ is open in $G$ because for any $g \in \langle V\rangle$, the set $g \cdot V = \{ga \mid a \in V\}$ is a neighborhood of $g$ in $G$ that is contained in $\langle V\rangle$. Second, to prove that $\langle V\rangle$ is closed in $G$, let $g \in G$ be a limit point of $\langle V\rangle$. The neighborhood $g \cdot V$ of $g$ in $G$ must contain some

$b \in \langle V \rangle$; that is, $ga = b$ for some $a, b \in \langle V \rangle$. So $g = ba^{-1}$ is a product of two elements of $\langle V \rangle$, which shows that $g \in \langle V \rangle$.

In summary, $\exp : \mathfrak{g} \to G$ is a surjective homomorphism whose kernel equals $\langle \{v_1, ..., v_k\} \rangle$. So,

$$G \cong \mathfrak{g}/\langle \{v_1, ..., v_k\} \rangle \cong T^k \times (\mathbb{R}^{d-k}, +),$$

where $d = \dim(\mathfrak{g})$. Since $G$ is compact, we must have $d = k$. $\square$

The above proof actually verifies the following more general theorem:

**Theorem 9.6.** *Any abelian path-connected matrix group is isomorphic to $T^k \times (\mathbb{R}^m, +)$ for some integers $k, m \geq 0$.*

## 2. The standard maximal torus and center of $SO(n)$, $SU(n)$, $U(n)$ and $Sp(n)$

**Definition 9.7.** *Let $G$ be a matrix group. A **torus in** $G$ means a subgroup of $G$ that is isomorphic to a torus. A **maximal torus in** $G$ means a torus in $G$ that is not contained in a higher-dimensional torus in $G$.*

Every matrix group $G$ contains at least one maximal torus, which is justified as follows. The subgroup $\{I\} \subset G$ is a 0-dimensional torus in $G$. If it is not contained in a 1-dimensional torus, then it is maximal. Otherwise, choose a 1-dimensional torus $T^1$ in $G$. If $T^1$ is not contained in a 2-dimensional torus, then it is maximal. Otherwise, choose a 2-dimensional torus $T^2$ in $G$ containing $T^1$, etc. This process must stop, since $G$ clearly cannot contain a torus with dimension higher than its own dimension.

Maximal tori are really only useful for studying *compact* matrix groups. So, in this section, we will determine maximal tori of our familiar compact matrix groups: $SO(n)$, $U(n)$, $SU(n)$ and $Sp(n)$.

We will use "diag" as a shorthand for **block-diagonal** matrices as well as diagonal matrices. For example,

$$\mathrm{diag}\left(\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}, \begin{pmatrix}5 & 6 & 7\\ 8 & 9 & 10\\ 11 & 12 & 13\end{pmatrix}, 14\right) = \begin{pmatrix}1 & 2 & 0 & 0 & 0 & 0\\ 3 & 4 & 0 & 0 & 0 & 0\\ 0 & 0 & 5 & 6 & 7 & 0\\ 0 & 0 & 8 & 9 & 10 & 0\\ 0 & 0 & 11 & 12 & 13 & 0\\ 0 & 0 & 0 & 0 & 0 & 14\end{pmatrix}.$$

Notice that the product of two similarly-shaped block-diagonal matrices is calculated blockwise. For example, when $A_1, B_1 \in M_{n_1}(\mathbb{K})$ and $A_2, B_2 \in M_{n_2}(\mathbb{K})$, we have:

$$\mathrm{diag}(A_1, A_2) \cdot \mathrm{diag}(B_1, B_2) = \mathrm{diag}(A_1 \cdot B_1, A_2 \cdot B_2) \in M_{n_1+n_2}(\mathbb{K}).$$

Therefore, if $G_1 \subset GL_{n_1}(\mathbb{K})$ and $G_2 \subset GL_{n_2}(\mathbb{K})$ are both matrix groups, then their product, $G_1 \times G_2$, is isomorphic to the following matrix group:

$$G_1 \times G_2 \cong \{\mathrm{diag}(A_1, A_2) \mid A_1 \in G_1, A_2 \in G_2\} \subset GL_{n_1+n_2}(\mathbb{K}).$$

Also notice that the determinant of a block-diagonal matrix is the product of the determinants of its blocks.

We also introduce notation for the familiar 2-by-2 rotation matrix:

$$\mathcal{R}_\theta = \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}.$$

**Theorem 9.8.** *Each of the following is a maximal torus.*

$$\begin{aligned}
T &= \{\mathit{diag}(\mathcal{R}_{\theta_1}, ..., \mathcal{R}_{\theta_m}) \mid \theta_i \in [0, 2\pi)\} \subset SO(2m).\\
T &= \{\mathit{diag}(\mathcal{R}_{\theta_1}, ..., \mathcal{R}_{\theta_m}, 1) \mid \theta_i \in [0, 2\pi)\} \subset SO(2m+1).\\
T &= \{\mathit{diag}(e^{\mathbf{i}\theta_1}, ..., e^{\mathbf{i}\theta_n}) \mid \theta_i \in [0, 2\pi)\} \subset U(n).\\
T &= \{\mathit{diag}(e^{\mathbf{i}\theta_1}, ..., e^{\mathbf{i}\theta_n}) \mid \theta_i \in [0, 2\pi)\} \subset Sp(n).\\
T &= \{\mathit{diag}(e^{\mathbf{i}\theta_1}, ..., e^{\mathbf{i}\theta_{n-1}}, e^{-\mathbf{i}(\theta_1+\cdots+\theta_{n-1})}) \mid \theta_i \in [0, 2\pi)\} \subset SU(n).
\end{aligned}$$

In each case, the given torus $T$ is called the **standard maximal torus** of the matrix group. It is not the only maximal torus, as we will see, but it is the simplest to describe. Notice that the standard maximal torus of $SU(n)$ is the intersection with $SU(n)$ of the standard maximal torus of $U(n)$.

**Proof.** In each case, it is easy to see that $T$ is a torus. The challenge is to prove that $T$ is not contained in a higher-dimensional torus of the group $G$. In each case, we will justify this by proving that any element $g \in G$ that commutes with all elements of $T$ must lie in $T$. Since any element of an alleged higher-dimensional torus would commute with all elements of $T$, this shows that no such higher-dimensional torus could exist.

CASE 1: $SO(n)$. For clarity, we will prove that

$$T = \{\text{diag}(\mathcal{R}_{\theta_1}, \mathcal{R}_{\theta_2}, 1) \mid \theta_1, \theta_2 \in [0, 2\pi)\}$$

is a maximal torus of $SO(5)$. Our arguments will generalize in an obvious way to $SO(n)$ for all even or odd $n$.

Suppose that $g \in SO(5)$ commutes with every element of $T$. Let $\theta$ be an angle that is not an integer multiple of $\pi$. We will use that $g$ commutes with $A = \text{diag}(\mathcal{R}_\theta, \mathcal{R}_\theta, 1) \in T$. Notice that real multiples of $e_5 = (0, 0, 0, 0, 1)$ are the only vectors in $\mathbb{R}^5$ that are fixed by $R_A$. Since $e_5 gA = e_5 Ag = e_5 g$, we learn that $R_A$ fixes $e_5 g$, which means $e_5 g = \pm e_5$. That is, the fifth row of $g$ looks like $(0, 0, 0, 0, \pm 1)$.

Next, use that $g$ commutes with $A = \text{diag}(\mathcal{R}_\theta, 1, 1, 1) \in T$. The only vectors in $\mathbb{R}^5$ fixed by $R_A$ are in $\text{span}\{e_3, e_4, e_5\}$. For each of $i \in \{3, 4\}$, we have that $e_i gA = e_i Ag = e_i g$, so $R_A$ fixes $e_i g$, which means that $e_i g \in \text{span}\{e_3, e_4, e_5\}$. In fact, $e_i g \in \text{span}\{e_3, e_4\}$, since otherwise $\langle e_i g, e_5 g\rangle \neq 0$. So the third and fourth rows of $g$ each has the form $(0, 0, a, b, 0)$. Repeating this argument with $A = \text{diag}(1, 1, \mathcal{R}_\theta, 1)$ gives that the first and second rows of $g$ each has the form $(a, b, 0, 0, 0)$.

In summary, $g$ has the form $g = \text{diag}(g_1, g_2, \pm 1)$ for some elements $g_1, g_2 \in M_2(\mathbb{R})$. Since $g \in SO(5)$, we have $g_1, g_2 \in O(2)$. It remains to prove that $g_1, g_2 \in SO(2)$, which forces the last argument to be $+1$ rather than $-1$ because $\det(g) = 1$. Suppose to the contrary that $g_1 \in O(2) - SO(2)$. Then $g$ does not commute with $\text{diag}(\mathcal{R}_\theta, 1, 1, 1)$. This is because $g_1$ does not commute with $\mathcal{R}_\theta$ by Exercise 3.6, which states that flips of $\mathbb{R}^2$ never commute with rotations of $\mathbb{R}^2$. Therefore $g_1 \in SO(2)$, and similarly $g_2 \in SO(2)$. Therefore, $g \in T$.

CASE 2: $U(n)$. We will prove that

$$T = \{\mathrm{diag}(e^{\mathbf{i}\theta_1}, ..., e^{\mathbf{i}\theta_n}) \mid \theta_i \in [0, 2\pi)\}$$

is a maximal torus of $U(n)$. Suppose that $g \in U(n)$ commutes with every element of $T$. Let $\theta$ be an angle that is not an integer multiple of $\pi$. We use that $g$ commutes with $A = \mathrm{diag}(e^{\mathbf{i}\theta}, e^{\mathbf{i}\theta}, ..., e^{\mathbf{i}\theta}, 1) \in T$. Notice that complex multiples of $e_n = (0, ..., 0, 1)$ are the only vectors in $\mathbb{C}^n$ fixed by $R_A$. Since $e_n gA = e_n Ag = e_n g$, we learn that $R_A$ fixes $e_n g$, which means $e_n g = \lambda e_n$ for some $\lambda \in \mathbb{C}$. That is, the $n^{th}$ row of $g$ looks like $(0, ..., 0, \lambda)$. Repeating this argument with the "1" entry of $A$ moved to other positions gives that $g$ is diagonal. It follows that $g \in T$.

CASE 3: $Sp(n)$. Suppose $g \in Sp(n)$ commutes with every element of $T = \{\mathrm{diag}(e^{\mathbf{i}\theta_1}, ..., e^{\mathbf{i}\theta_n}) \mid \theta_i \in [0, 2\pi)\}$. The argument in case 2 gives that $g$ is diagonal; that is, $g = \mathrm{diag}(q_1, ..., q_n)$ for some $q_i \in \mathbb{H}$. Since $g$ commutes with $\mathrm{diag}(\mathbf{i}, 1, ..., 1) \in T$, we know that $q_1\mathbf{i} = \mathbf{i}q_1$. By Exercise 1.15, this implies that $q_1 \in \mathbb{C}$. Similarly, $q_i \in \mathbb{C}$ for $i = 1, ..., n$. It follows that $g \in T$.

CASE 4: $SU(n)$. For clarity, we will prove that

$$T = \{\mathrm{diag}(e^{\mathbf{i}\theta_1}, e^{\mathbf{i}\theta_2}, e^{-\mathbf{i}(\theta_1+\theta_2)}) \mid \theta_1, \theta_2 \in [0, 2\pi)\}$$

is a maximal torus of $SU(3)$. Suppose that $g \in SU(3)$ commutes with every element of $T$. Since $g$ commutes with $A = \mathrm{diag}(1, e^{\mathbf{i}\theta}, e^{-\mathbf{i}\theta}) \in T$, the first row of $g$ must be a multiple of $e_1$. Permuting the three diagonal entries of $A$ gives that the second row of $g$ is a multiple of $e_2$ and the third of $e_3$. So $g$ is diagonal, which implies that $g \in T$.

This argument generalizes to $SU(n)$ for all $n \geq 3$. It remains to prove that $T = \{\mathrm{diag}(e^{\mathbf{i}\theta}, e^{-\mathbf{i}\theta}) \mid \theta \in [0, 2\pi)\}$ is a maximal torus of $SU(2)$. The isomorphism from $Sp(1)$ to $SU(2)$ (Section 4 of Chapter 3) sends the standard maximal torus of $Sp(1)$ to $T$, so this follows from case 3. $\square$

In each case of the previous proof, we verified the maximality of the standard torus by proving something slightly stronger:

**Proposition 9.9.** *Let $G \in \{SO(n), U(n), SU(n), Sp(n)\}$, and let $T$ be the standard maximal torus of $G$. Then any element of $G$ that commutes with every element of $T$ must lie in $T$. In particular, $T$ is* ***maximal abelian****, which means that $T$ is not contained in any larger abelian subgroup of $G$.*

As an application, we will calculate the centers of $SO(n)$, $U(n)$, $Sp(n)$ and $SU(n)$. Remember that the **center** of a group $G$ is defined as

$$Z(G) = \{g \in G \mid ga = ag \text{ for all } a \in G\}.$$

**Theorem 9.10.**

(1) $Z(SO(2m)) = \{I, -I\}$ *(the group of order 2) if $m > 1$.*
(2) $Z(SO(2m+1)) = \{I\}$ *(the trivial group).*
(3) $Z(U(n)) = \{e^{\mathbf{i}\theta} \cdot I \mid \theta \in [0, 2\pi)\}$ *(isomorphic to $U(1)$).*
(4) $Z(Sp(n)) = \{I, -I\}$.
(5) $Z(SU(n)) = \{\omega \cdot I \mid \omega^n = 1\}$ *(the cyclic group of order $n$).*

Notice that $Z(SU(n)) = Z(U(n)) \cap SU(n)$.

**Proof.** By Proposition 9.9, the center of each of these groups is a subset of its standard maximal torus. From this starting point, the arguments are straightforward, so we will leave to the reader all but the case $G = U(n)$.

Suppose that $g \in Z(U(n))$. Since $g$ lies in the standard maximal torus, it must be diagonal: $g = \text{diag}(\lambda_1, ..., \lambda_n)$. We will use that $g$ commutes with $A = \text{diag}\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, 1, ..., 1\right) \in U(n)$:

$$\begin{aligned} \text{diag}\left(\begin{pmatrix} 0 & \lambda_1 \\ \lambda_2 & 0 \end{pmatrix}, \lambda_3, ..., \lambda_n\right) &= gA \\ &= Ag = \text{diag}\left(\begin{pmatrix} 0 & \lambda_2 \\ \lambda_1 & 0 \end{pmatrix}, \lambda_3, ..., \lambda_n\right), \end{aligned}$$

which implies that $\lambda_1 = \lambda_2$. By a similar argument, any other pair of $\lambda$'s must be equal. So $g$ has the form $\text{diag}(\lambda, ..., \lambda) = \lambda \cdot I$ for some $\lambda \in \mathbb{C}$ with unit norm. $\square$

**Corollary 9.11.**

(1) *$SU(2)$ is not isomorphic to $SO(3)$.*
(2) *$SU(n)\times U(1)$ is not isomorphic to $U(n)$ (for $n>1$).*

**Proof.** Their centers are not isomorphic. □

Remember that $SU(2)$ and $SO(3)$ have isomorphic Lie algebras. There exists a 2-to-1 homomorphism from $SU(2)$ to $SO(3)$ which is a local diffeomorphism. Corollary 9.11 (or the fact that they are not homeomorphic by Proposition 8.20) says that "2-to-1" cannot be improved to "1-to-1".

The pair $SU(n)\times U(1)$ and $U(n)$ are diffeomorphic, but the natural diffeomorphism between them does not preserve the group structure. They have isomorphic Lie algebras because there is an $n$-to-1 homomorphism from $SU(n)\times U(1)$ to $U(n)$ which is a local diffeomorphism. These statements are all justified in Exercise 4.21. The corollary implies that "$n$-to-1" cannot be improved to "1-to-1".

## 3. Conjugates of a maximal torus

The standard maximal tori are not the only maximal tori of $SO(n)$, $U(n)$, $Sp(n)$ and $SU(n)$. Other ones are obtained by conjugating the standard ones.

**Proposition 9.12.** *If $T$ is a maximal torus of a matrix group $G$, then for any $g\in G$, $gTg^{-1}=\{gag^{-1}\mid a\in T\}$ is also a maximal torus of $G$.*

**Proof.** The conjugation map $C_g: G\to G$, which sends $a\mapsto gag^{-1}$, is an isomorphism. So the image of $T$ under $C_g$, namely $gTg^{-1}$, is isomorphic to $T$ and is therefore a torus. If $\tilde{T}\subset G$ were a higher-dimensional torus containing $gTg^{-1}$, then $C_g^{-1}(\tilde{T})$ would be a higher-dimensional torus containing $T$. This is not possible, so $gTg^{-1}$ must be maximal. □

Since a maximal torus is not typically a normal subgroup, it differs from some of its conjugates. The main result of this section is that there are enough different conjugates to cover the whole group.

**Theorem 9.13.** *Let $G \in \{SO(n), U(n), SU(n), Sp(n)\}$, and let $T$ be the standard maximal torus of $G$. Then any element of $G$ is contained in $gTg^{-1}$ for some $g \in G$.*

A more general fact is true, which we will not prove: the conjugates of any maximal torus of any path-connected compact matrix group cover the group.

Theorem 9.13 says that:

$$\text{For each } x \in G, \text{ there exists } g \in G \text{ such that } x \in gTg^{-1}. \tag{9.1}$$

This is equivalent to:

$$\text{For each } x \in G, \text{ there exists } g \in G \text{ such that } gxg^{-1} \in T. \tag{9.2}$$

In other words, every $x \in G$ can be conjugated into the diagonal or block-diagonal form that characterizes elements of the standard maximal torus.

In Equation 9.2, think of $g$ as a change of basis matrix, as explained in Section 7 of Chapter 1. The linear transformation $R_x$ is represented with respect to the orthonormal basis $\{e_1g, ..., e_ng\}$ by the matrix $gxg^{-1} \in T$.

The example $G = SO(3)$ helps clarify this idea. Let $x \in SO(3)$. The theorem insures that there exists $g \in SO(3)$ such that

$$gxg^{-1} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

for some $\theta \in [0, 2\pi)$. This is the matrix representing $R_x$ in the basis $\{e_1g, e_2g, e_3g\}$, which means that $R_x$ is a rotation through angle $\theta$ about the line spanned by $e_3g$. To verify this explicitly, notice:

$$\begin{aligned} e_3(gxg^{-1}) = e_3 &\Rightarrow (e_3g)x = e_3g, \\ e_1(gxg^{-1}) = (\cos\theta)e_1 + (\sin\theta)e_2 &\Rightarrow (e_1g)x = (\cos\theta)e_1g + (\sin\theta)e_2g, \\ e_2(gxg^{-1}) = (\cos\theta)e_2 - (\sin\theta)e_1 &\Rightarrow (e_2g)x = (\cos\theta)e_2g - (\sin\theta)e_1g. \end{aligned} \tag{9.3}$$

We conclude that *every element of $SO(3)$ represents a rotation*!

Analogous interpretations hold for $SO(n)$. Take $SO(5)$ for example. An element, $y$, of the standard maximal torus of $SO(5)$ is particularly simple: $R_y$ represents a rotation by some angle $\theta_1$ in the plane span$\{e_1, e_2\}$ and a simultaneous rotation by a second angle $\theta_2$ in the plane span$\{e_3, e_4\}$. The theorem says that every $x \in SO(5)$ is equally simple. There exist $g \in SO(5)$ such that $R_x$ represents a simultaneous rotation in the planes span$\{e_1 g, e_2 g\}$ and span$\{e_3 g, e_4 g\}$. Notice that these two planes are orthogonal because $g \in SO(5)$. Similarly, every element of $SO(n)$ represents a simultaneous rotation in a collection of orthogonal planes.

Before proving Theorem 9.13, we review some linear algebra terminology. Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, and let $f : \mathbb{K}^n \to \mathbb{K}^n$ be a linear transformation. Recall that $\lambda \in \mathbb{K}$ is called an **eigenvalue** of $f$ if $f(v) = \lambda \cdot v$ for some non-zero $v \in \mathbb{K}^n$. Notice that for any $\lambda \in \mathbb{K}$,

$$\mathcal{V}(\lambda) = \{v \in \mathbb{K}^n \mid f(v) = \lambda \cdot v\}$$

is a subspace of $\mathbb{K}^n$ (this is false for $\mathbb{K} = \mathbb{H}$). Notice that $\lambda$ is an eigenvalue of $f$ exactly when $\mathcal{V}(\lambda)$ has dimension $\geq 1$. The non-zero vectors in $\mathcal{V}(\lambda)$ are called **eigenvectors** associated to $\lambda$. For a matrix $A \in M_n(\mathbb{K})$, a basic fact from linear algebra is: $\lambda \in \mathbb{K}$ is an eigenvalue of $R_A$ if and only if $\det(A - \lambda \cdot I) = 0$.

**Lemma 9.14.** *Any linear transformation $f : \mathbb{C}^n \to \mathbb{C}^n$ has an eigenvalue.*

**Proof.** $f = R_A$ for some $A \in M_n(\mathbb{C})$. The fundamental theorem of algebra says that every polynomial of degree $\geq 1$ with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$. In particular,

$$g(\lambda) = \det(A - \lambda \cdot I)$$

equals zero for some $\lambda \in \mathbb{C}$, which is an eigenvalue of $f$. $\square$

**Proposition 9.15.** *For every $A \in U(n)$, $R_A$ has an orthonormal basis of eigenvectors.*

**Proof.** Let $A \in U(n)$. By Lemma 9.14, there exists an eigenvalue $\lambda_1$ of $R_A$. Let $v_1 \in \mathbb{C}^n$ be an associated eigenvector, which can be chosen to have norm 1, since any multiple of an eigenvector is an eigenvector. Notice that $\lambda_1 \neq 0$, since $R_A$ is invertible.

To find a second eigenvector, we use that $A$ is unitary. The key observation is: if $w \in \mathbb{C}^n$ is orthogonal to $v_1$ (with respect to the hermitian inner product), then $R_A(w)$ is also orthogonal to $v_1$. To justify this, notice that:

$$\langle wA, v_1\rangle = \langle w, v_1 A^{-1}\rangle = \left\langle w, \frac{1}{\lambda_1} v_1\right\rangle = \langle w, v_1\rangle(1/\overline{\lambda}_1) = 0. \tag{9.4}$$

This means that $R_A : \mathbb{C}^n \to \mathbb{C}^n$ restricts to a linear transformation,

$$R_A : \text{span}\{v_1\}^\perp \to \text{span}\{v_1\}^\perp,$$

where

$$\text{span}\{v_1\}^\perp = \{w \in \mathbb{C}^n \mid w \text{ is orthogonal to } v_1\},$$

which is an $(n-1)$-dimensional $\mathbb{C}$-subspace of $\mathbb{C}^n$. By applying Lemma 9.14 a second time, the restricted $R_A$ has a unit-length eigenvector $v_2 \in \text{span}\{v_1\}^\perp$, and $R_A$ restricts further to a linear transformation

$$R_A : \text{span}\{v_1, v_2\}^\perp \to \text{span}\{v_1, v_2\}^\perp.$$

Repeating this argument a total of $n$ times proves the lemma. □

As a corollary of Proposition 9.15, we prove that Theorem 9.13 is true when $G = U(n)$.

**Corollary 9.16.** *For any $A \in U(n)$, there exists $g \in U(n)$ such that $gAg^{-1}$ is diagonal and hence lies in the standard maximal torus of $U(n)$.*

**Proof.** Let $\{v_1, ..., v_n\}$ be an orthonormal basis of eigenvectors of $R_A$ with eigenvalues $\{\lambda_1, ..., \lambda_n\}$. Let $g$ denote the matrix whose $i^{\text{th}}$ row equals $v_i$, so that $e_i g = v_i$, for each $i = 1, ..., n$. Notice that $g \in U(n)$. We claim that

$$gAg^{-1} = \text{diag}(\lambda_1, ..., \lambda_n).$$

This is simply because $gAg^{-1}$ represents the linear transformation $R_A$ in the basis $\{v_1, ..., v_n\}$. To understand this more concretely, notice that for each $i = 1, ..., n$,

$$e_i(gAg^{-1}) = v_i A g^{-1} = \lambda_i v_i g^{-1} = \lambda_i e_i.$$

□

Next we verify Theorem 9.13 when $G = SU(n)$.

**Corollary 9.17.** *For any $A \in SU(n)$, there exists $g \in SU(n)$ such that $gAg^{-1}$ is diagonal and hence lies in the standard maximal torus of $SU(n)$.*

**Proof.** Let $A \in SU(n)$. By the previous corollary, there exists some $g \in U(n)$ such that $gAg^{-1}$ is diagonal. Notice that for any $\theta \in [0, 2\pi)$,

$$(e^{i\theta}g)A(e^{i\theta}g)^{-1} = gAg^{-1}.$$

Further, $\theta$ can easily be chosen such that $e^{i\theta}g \in SU(n)$. □

The $U(n)$ case also helps us prove the $Sp(n)$ case:

**Corollary 9.18.** *For any $A \in Sp(n)$, there exists $g \in Sp(n)$ such that $gAg^{-1}$ is diagonal with all entries in $\mathbb{C}$ and hence lies in the standard maximal torus of $Sp(n)$.*

**Proof.** Let $A \in Sp(n)$. Recall from Chapter 2 that the injective homomorphism $\Psi_n : Sp(n) \to U(2n)$ is defined such that the following diagram commutes:

$$\begin{tikzcd} \mathbb{H}^n \arrow[r, "g_n"] \arrow[d, "R_A"'] & \mathbb{C}^{2n} \arrow[d, "R_{\Psi_n(A)}"] \\ \mathbb{H}^n \arrow[r, "g_n"] & \mathbb{C}^{2n} \end{tikzcd}$$

Since every unitary matrix has a unit-length eigenvalue, there exists $u_1 \in \mathbb{C}^{2n}$ such that $R_{\Psi_n(A)}(u_1) = \lambda_1 u_1$ for some $\lambda_1 \in \mathbb{C}$ with $|\lambda_1| = 1$. Let $v_1 = g_n^{-1}(u_1) \in \mathbb{H}^n$. We claim that $R_A(v_1) = \lambda_1 v_1$. This is because:

$$g_n(\lambda_1 v_1) = \lambda_1 g_n(v_1) = \lambda_1 u_1 = R_{\Psi_n(A)}(u_1) = g_n(R_A(v_1)).$$

Next notice that if $w \in \mathbb{H}^n$ is orthogonal to $v_1$ (with respect to the symplectic inner product), then so is $R_A(w)$. The verification is identical to Equation 9.4. So $R_A : \mathbb{H}^n \to \mathbb{H}^n$ restricts to an $\mathbb{H}$-linear function from the following $(n-1)$-dimensional $\mathbb{H}$-subspace of $\mathbb{H}^n$ to itself:

$$\text{span}\{v_1\}^\perp = \{w \in \mathbb{H}^n \mid w \text{ is orthogonal to } v_1\}.$$

Therefore, $R_{\Psi_n(A)} : \mathbb{C}^{2n} \to \mathbb{C}^{2n}$ restricts to a linear function from $g_n(\text{span}\{v_1\}^\perp)$ to itself. Let $u_2 \in \mathbb{C}^{2n}$ be a unit-length eigenvector of this restriction of $R_{\Psi_n(A)}$, with eigenvalue $\lambda_2$, and let $v_2 = g_n^{-1}(u_2)$. As before, $R_A(v_2) = \lambda_2 v_2$. Repeating this argument a total of $n$ times produces an orthonormal basis $\{v_1, ..., v_n\}$ of $\mathbb{H}^n$ and unit-length complex numbers $\{\lambda_1, ..., \lambda_n\}$ such that $R_A(v_i) = \lambda_i v_i$ for each $i$.

Finally, if $g \in Sp(n)$ is the matrix whose rows are $v_1, ..., v_n$, then

$$gAg^{-1} = \text{diag}(\lambda_1, ..., \lambda_n),$$

exactly as in the proof of Corollary 9.16. $\square$

Finally, we prove Theorem 9.13 in the case $G = SO(n)$.

**Proposition 9.19.** *For any $A \in SO(n)$, there exists $g \in SO(n)$ such that $gAg^{-1}$ lies in the standard maximal torus of $SO(n)$.*

**Proof.** Let $A \in SO(n)$. We can regard $A$ as an $n$ by $n$ complex matrix whose entries happen to all be real numbers. Regarded as such, $A \in SU(n)$, so there exists $v \in \mathbb{C}^n$ such that $vA = \lambda v$ for some unit-length $\lambda \in \mathbb{C}$. Let $\overline{v} \in \mathbb{C}^n$ denote the result of conjugating all of the entries of $v$. Notice that:

$$\overline{v}A = \overline{v}\overline{A} = \overline{vA} = \overline{\lambda v} = \overline{\lambda}\overline{v}, \tag{9.5}$$

so $\overline{v}$ is also an eigenvector of $R_A$, with eigenvalue $\overline{\lambda}$.

**CASE 1**: Suppose $\lambda \in \mathbb{R}$ (so $\lambda = \overline{\lambda} = \pm 1$). In this case, Equation 9.5 says that $\overline{v}$ is also an eigenvector associated to $\lambda$. Thus, $\text{span}_{\mathbb{C}}\{v, \overline{v}\} \subset \mathcal{V}(\lambda) =$ the space of eigenvectors associated to $\lambda$. We claim that $\text{span}_{\mathbb{C}}\{v, \overline{v}\}$ contains a vector, $Z$, with all real components. Specifically, $Z = \frac{v+\overline{v}}{|v+\overline{v}|}$ works unless its denominator vanishes, in which case $Z = \frac{\mathbf{i}V}{|\mathbf{i}V|}$ works.

**CASE 2**: Suppose $\lambda \notin \mathbb{R}$. Write $\lambda = e^{\mathbf{i}\theta}$ for some angle $\theta$, which is not an integer multiple of $\pi$. Define:

$$X = v + \overline{v},$$
$$Y = \mathbf{i}(v - \overline{v}).$$

All entries of $X$ and $Y$ are real, so $X, Y \in \mathbb{R}^n$. It is straightforward to check that $X$ and $Y$ are orthogonal. Observe that:

$$\begin{aligned}(9.6)\qquad XA &= (v+\overline{v})A = e^{\mathbf{i}\theta}v + e^{-\mathbf{i}\theta}\overline{v}\\ &= (\cos\theta + \mathbf{i}\sin\theta)v + (\cos\theta - \mathbf{i}\sin\theta)\overline{v}\\ &= (\cos\theta)(v+\overline{v}) + (\sin\theta)(\mathbf{i}v - \mathbf{i}\overline{v})\\ &= (\cos\theta)X + (\sin\theta)Y.\\ \text{Similarly, } YA &= (-\sin\theta)X + (\cos\theta)Y.\end{aligned}$$

Using the fact that $\theta$ is not a multiple of $\pi$, Equation 9.6 implies that $X$ and $Y$ have the same norm, which is non-zero since $v \neq 0$. So $R_A$ rotates $\text{span}\{X, Y\} \subset \mathbb{R}^n$ by an angle $\theta$. If $X$ and $Y$ are re-scaled to have unit-length, they still satisfy the punchline of Equation 9.6:

$$XA = (\cos\theta)X + (\sin\theta)Y,$$
$$YA = (-\sin\theta)X + (\cos\theta)Y.$$

In case 1, let $\Omega = \text{span}(Z) \subset \mathbb{R}^n$. In case 2, let

$$\Omega = \text{span}(X, Y) \subset \mathbb{R}^n.$$

In either case, $\Omega$ is stable under $R_A : \mathbb{R}^n \to \mathbb{R}^n$, meaning that $R_A(w) \in \Omega$ for all $w \in \Omega$. By an argument analogous to Equation 9.4, the subspace

$$\Omega^\perp = \{w \in \mathbb{R}^n \mid w \text{ is orthogonal to every element of } \Omega\}$$

is also stable under $R_A$. So we can repeat the above argument on the restriction of $R_A$ to $\Omega^\perp$.

Repeating this argument enough times produces an orthonormal basis of $\mathbb{R}^n$ of the form $\{X_1, Y_1, ..., X_k, Y_k, Z_1, ..., Z_l\}$, with $2k+l = n$. If $g$ is the matrix whose rows equal these basis vectors, then

$$gAg^{-1} = \text{diag}(\mathcal{R}_{\theta_1}, ..., \mathcal{R}_{\theta_k}, \lambda_1, ..., \lambda_l),$$

where each $\lambda$ is $\pm 1$. By re-ordering the basis, we can assume that the negative lambda's come first. There are an even number of negative lambda's because $\det(gAg^{-1}) = \det(A) = 1$. Each pair of negative lambda's is a rotation block, since $\text{diag}(-1, -1) = \mathcal{R}_\pi$. It follows that $gAg^{-1}$ lies in the standard maximal torus $T$ of $SO(n)$.

Since the basis is orthonormal, $g \in O(n)$. If $g \in SO(n)$, then we are done, so assume that $g \in O(n) - SO(n)$. In this case, define:

$$a = \operatorname{diag}\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, 1, ..., 1\right) \in O(n) - SO(n).$$

Notice that $ag \in SO(n)$ and that $aTa^{-1} = T$, so

$$(ag)A(ag)^{-1} = a(gAg^{-1})a^{-1} \in T,$$

which verifies that $ag \in SO(n)$ conjugates $A$ into $T$. $\square$

This completes our proof of Theorem 9.13.

## 4. The Lie algebra of a maximal torus

In this section, let $G \in \{SO(n), U(n), SU(n), Sp(n)\}$ and let $\mathfrak{g}$ be the Lie algebra of $G$. Let $T = T(G) \subset G$ be the standard maximal torus of $G$, and let $\tau = \tau(\mathfrak{g}) \subset \mathfrak{g}$ be the Lie algebra of $T$. It is straightforward to calculate:

$$\tau(so(2m)) = \left\{\operatorname{diag}\left(\begin{pmatrix} 0 & \theta_1 \\ -\theta_1 & 0 \end{pmatrix}, ..., \begin{pmatrix} 0 & \theta_m \\ -\theta_m & 0 \end{pmatrix}\right) \mid \theta_i \in \mathbb{R}\right\},$$

$$\tau(so(2m+1)) = \left\{\operatorname{diag}\left(\begin{pmatrix} 0 & \theta_1 \\ -\theta_1 & 0 \end{pmatrix}, ..., \begin{pmatrix} 0 & \theta_m \\ -\theta_m & 0 \end{pmatrix}, 0\right) \mid \theta_i \in \mathbb{R}\right\},$$

(9.7)

$$\begin{aligned} \tau(u(n)) &= \{\operatorname{diag}(\mathbf{i}\theta_1, ..., \mathbf{i}\theta_n) \mid \theta_i \in \mathbb{R}\}, \\ \tau(sp(n)) &= \{\operatorname{diag}(\mathbf{i}\theta_1, ..., \mathbf{i}\theta_n) \mid \theta_i \in \mathbb{R}\}, \\ \tau(su(n)) &= \{\operatorname{diag}(\mathbf{i}\theta_1, ..., \mathbf{i}\theta_{n-1}, -\mathbf{i}(\theta_1 + \cdots + \theta_{n-1})) \mid \theta_i \in \mathbb{R}\}. \end{aligned}$$

Compare to Theorem 9.8, where we described $T(G)$ using the same parameters $\theta_i$ that are used above to describe $\tau(\mathfrak{g})$. The descriptions correspond via matrix exponentiation. The exponential image of a vector in $\tau(\mathfrak{g})$ equals the element of $T(G)$ described by the same angles. In $U(n)$ for example,

$$e^{\operatorname{diag}(\mathbf{i}\theta_1, ..., \mathbf{i}\theta_n)} = \operatorname{diag}(e^{\mathbf{i}\theta_1}, ..., e^{\mathbf{i}\theta_n}) \in T(U(n)).$$

Since $T$ is abelian, its Lie algebra, $\tau$, is abelian, which means that the Lie bracket of any pair of matrices in $\tau$ equals zero.

Using the fact that all elements of $G$ can be conjugated into $T$, we will show that all elements of $\mathfrak{g}$ can be conjugated into $\tau$.

**Proposition 9.20.** *For each $X \in \mathfrak{g}$ there exists $g \in G$ such that $Ad_g(X) \in \tau$.*

**Proof.** Choose $r > 0$ such that $\exp : \mathfrak{g} \to G$ is a diffeomorphism on the ball in $\mathfrak{g}$ of radius $r$ centered at the origin. It will suffice to prove the proposition for $X \in \mathfrak{g}$ with $|X| < r$. By Theorem 9.13, there exists $g \in G$ such that $a = g(e^X)g^{-1} \in T$, so:

$$e^{\mathrm{Ad}_g(X)} = e^{gXg^{-1}} = g(e^X)g^{-1} = a \in T.$$

Remember that $|\mathrm{Ad}_g(X)| = |X| < r$, so $\mathrm{Ad}_g(X)$ is the unique vector with length $< r$ that exponentiates to $a \in T$. Equation 9.7 explicitly describes this vector in terms of the angles $\theta_i$ of $a$; in particular it lies in $\tau$. $\square$

Proposition 9.20 is important in linear algebra. It says that any skew-symmetric or skew-hermitian or skew-symplectic matrix can be conjugated into the diagonal or block-diagonal form of Equation 9.7. This adds to the list in Theorem 9.13 of matrix types that can be conjugated into simple forms. In fact, Theorem 9.13 and Proposition 9.20 together give a beautifully uniform way of understanding many conjugation theorems from linear algebra!

A key application of Theorem 9.13 is the following proposition, which implies in particular that $G$ is path-connected:

**Proposition 9.21.** *The exponential map* $\exp : \mathfrak{g} \to G$ *is surjective.*

**Proof.** We have an explicit description of the restriction $\exp : \tau \to T$, which is clearly surjective. For any $g \in G$, $gTg^{-1}$ is a maximal torus with Lie algebra $Ad_g(\tau)$. Also, the restriction $\exp : \mathrm{Ad}_g(\tau) \to gTg^{-1}$ is surjective, since $e^{\mathrm{Ad}_g(X)} = ge^X g^{-1}$ for all $X \in \tau$. Theorem 9.13 says that these conjugates cover $G$. $\square$

## 5. The shape of $SO(3)$

In Section 6 of Chapter 8, we identified $SO(3)$ with $\mathbb{RP}^3$. We will now give a different identification, which relies on explicitly understanding the exponential map $\exp : so(3) \to SO(3)$.

Recall from Section 3 of Chapter 8 the following vector space isomorphism $f : \mathbb{R}^3 \to so(3)$:

$$(a, b, c) \overset{f}{\mapsto} \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}.$$

Recall that under this identification, $\mathrm{Ad}_g : so(3) \to so(3)$ corresponds to $L_g : \mathbb{R}^3 \to \mathbb{R}^3$ for all $g \in SO(3)$. More precisely,

$$\mathrm{Ad}_g(f(a, b, c)) = f(L_g(a, b, c)). \tag{9.8}$$

**Proposition 9.22.** *For any $A \in so(3)$, $L_{(e^A)} : \mathbb{R}^3 \to \mathbb{R}^3$ is a right-handed rotation through angle $|A|/\sqrt{2}$ about the axis spanned by $f^{-1}(A)$.*

"Right-handed" means the rotation is in the direction that the fingers of your right hand curl when your thumb is pointed towards $f^{-1}(A)$.

**Proof.** Let $A \in so(3)$. By Proposition 9.20, there exists $g \in SO(3)$ such that

$$A = g \begin{pmatrix} 0 & -\theta & 0 \\ \theta & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} g^{-1}$$

for some $\theta \in \mathbb{R}$. Notice that $|A| = \sqrt{2}\theta$, so $\theta = |A|/\sqrt{2}$. Next,

$$\exp(A) = g \exp \begin{pmatrix} 0 & -\theta & 0 \\ \theta & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} g^{-1} = g \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} g^{-1}.$$

It follows that $L_{(e^A)}$ is a right-handed rotation through angle $\theta$ about the line spanned by $g \cdot (0, 0, 1)$. The verification is similar to Equation 9.3.

Finally, notice that by Equation 9.8, the rotation axis is:

$$g \cdot (0, 0, 1) = f^{-1} \left( \mathrm{Ad}_g \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right) = \frac{1}{\theta} f^{-1}(A).$$

$\square$

**Corollary 9.23.** *Let $B = \{A \in so(3) \mid |A| \leq \pi\sqrt{2}\}$. The restriction $\exp : B \to SO(3)$ is surjective. It is not injective, but for $A_1, A_2 \in B$, $\exp(A_1) = \exp(A_2)$ if and only if $A_1 = \pm A_2$ and $|A_1| = |A_2| = \pi\sqrt{2}$.*

**Proof.** The image $\exp(B)$ contains matrices representing all right-handed rotations about all vectors in $\mathbb{R}^3$ through all angles $\theta \in [0, \pi]$. Notice that the right-handed rotation through angle $\theta$ about $A \in \mathbb{R}^3$ equals the right-handed rotation through angle $-\theta$ about $-A$. This is why $\exp(A) = \exp(-A)$ when $|A| = \pi\sqrt{2}$. □

Points of $SO(3)$ are in one-to-one correspondence with points of $B/\sim$, where $\sim$ is the equivalence relationship on $B$ that identified each point on the boundary of $B$ with its antipode (its negative).

What does this have to do with $\mathbb{RP}^3$? Well, $B$ is homeomorphic to the "upper-hemisphere," $V$, of $S^3$:

$$V = \{(x_0, x_1, x_2, x_3) \in S^3 \subset \mathbb{R}^4 \mid x_0 \geq 0\},$$

by an argument analogous to the proof of Proposition 7.15. A typical line through the origin in $\mathbb{R}^4$ intersects $V$ in exactly one point. The only exceptions are the lines in the subspace $\{x_0 = 0\}$; these intersect $V$ in a pair of antipodal points on its boundary. So $\mathbb{RP}^3$ can be modelled as the upper hemisphere $V$ modulo identification of antipodal boundary pairs. This is another way of understanding why $SO(3)$ is naturally identified with $\mathbb{RP}^3$.

## 6. The rank of a compact matrix group

Let $G \in \{SO(n), U(n), SU(n), Sp(n)\}$. Let $T$ be the standard maximal torus of $G$. In this section we prove the following:

**Theorem 9.24.** *Every maximal torus of $G$ equals $gTg^{-1}$ for some $g \in G$.*

Our proof actually holds when $T$ is any maximal torus of any path-connected compact matrix group, granting the previously mentioned fact that the conjugates of $T$ cover $G$ in this generality.

In particular, any two maximal tori of $G$ have the same dimension, so the following is well-defined:

**Definition 9.25.** *The* ***rank*** *of $G$ is the dimension of a maximal torus.*

The ranks of our familiar compact groups are:

$$\begin{aligned}\operatorname{rank}(SO(2n)) = \operatorname{rank}(SO(2n+1)) = \operatorname{rank}(U(n))\\ = \operatorname{rank}(Sp(n)) = \operatorname{rank}(SU(n+1)) = n.\end{aligned}$$

Isomorphic groups clearly have the same rank, so rank is a useful invariant for proving that two groups are not isomorphic. The proof of Theorem 9.24 relies on a useful fact about tori:

**Lemma 9.26.** *For any $n$, there exists $a \in T^n$ such that the set $\{a, a^2, a^3, a^4, ...\}$ is dense in $T^n$.*

**Proof of the $n = 1$ case.** Let $\theta$ be an irrational angle, which means an irrational multiple of $\pi$. Let $a = (e^{\mathbf{i}\theta}) \in T^1 = U(1)$. To verify that this choice works, notice that $\{a = (e^{\mathbf{i}\theta}), a^2 = (e^{2\mathbf{i}\theta}), a^3 = (e^{3\mathbf{i}\theta}), ...\}$ is an infinite sequence of points that are all *distinct* because $\theta$ is irrational. Since $U(1)$ is compact, some subsequence converges (Proposition 4.23). This convergent subsequence must be Cauchy, which means that for any $\epsilon > 0$, we can find integers $n_1 < n_2$ such that $\operatorname{dist}(a^{n_1}, a^{n_2}) < \epsilon$. Next, notice that for any integer $m$,

$$\begin{aligned}\operatorname{dist}(a^{m+(n_2-n_1)}, a^m) &= \operatorname{dist}(a^m a^{n_2} a^{-n_1}, a^m)\\ &= \operatorname{dist}(a^{n_2} a^{-n_1}, I) = \operatorname{dist}(a^{n_2}, a^{n_1}) < \epsilon.\end{aligned}$$

So the sequence $\{a^{(n_2-n_1)}, a^{2(n_2-n_1)}, a^{3(n_2-n_1)}, ...\}$ takes baby steps of uniform size $< \epsilon$ and thus comes within a distance $\epsilon$ of every element of $U(1)$ as it hops around the circle. Since $\epsilon > 0$ was arbitrary, the lemma follows. □

For $n > 1$, we must choose $a = (e^{\mathbf{i}\theta_1}, ..., e^{\mathbf{i}\theta_n}) \in T^n$ such that the $\theta$'s are *rationally independent*, which means there are no equalities of the form $\sum_{k=1}^{n} s_k\theta_k = \pi$, where $s_k$ are rational numbers. The proof that this works is found in [**2**, page 66]. An alternative purely topological proof of Lemma 9.26 is found in [**1**].

**Proof of Theorem 9.24.** Let $T' \subset G$ be a maximal torus. Choose $a \in T'$ such that $\{a, a^2, a^3, ...\}$ is dense in $T'$. Choose $g \in G$ such that

$gag^{-1} \in T$. Since $T$ is a subgroup, $(gag^{-1})^n = ga^n g^{-1} \in T$ for every integer $n$, so a dense subset of $gT'g^{-1}$ lies in $T$. Since $T$ is closed, $gT'g^{-1} \subset T$. Since $gT'g^{-1}$ is a *maximal* torus, $gT'g^{-1} = T$. □

## 7. Exercises

**Ex. 9.1.** In Theorem 9.10, prove the remaining cases $G = SU(n)$, $G = SO(n)$, $G = Sp(n)$.

**Ex. 9.2.** Prove that the standard maximal torus of $SO(3)$ is also a maximal torus of $GL_3(\mathbb{R})$. Do its conjugates cover $GL_3(\mathbb{R})$?

**Ex. 9.3.** If $T_1 \subset G_1$ and $T_2 \subset G_2$ are maximal tori of matrix groups $G_1$ and $G_2$, prove that $T_1 \times T_2$ is a maximal torus of $G_1 \times G_2$.

**Ex. 9.4.** Prove $U(n)/Z(U(n))$ is isomorphic to $SU(n)/Z(SU(n))$.

**Ex. 9.5.** Use maximal tori to find a simple proof that if $A \in U(n)$, then

$$\det(e^A) = e^{\text{trace}(A)}.$$

This is a special case of Lemma 6.15.

**Ex. 9.6.** Let $A = \text{diag}(1, 1, ..., 1, -1) \in O(n)$. An element of $O(n)$ of the form $gAg^{-1}$ for $g \in O(n)$ is called a **reflection**.

(1) Show that $R_A : \mathbb{R}^n \to \mathbb{R}^n$ fixes $\text{span}\{e_1, ..., e_{n-1}\}$ and can be visualized as a reflection across this subspace.
(2) Show that $R_{gAg^{-1}} : \mathbb{R}^n \to \mathbb{R}^n$ fixes $\text{span}\{e_1 g^{-1}, ..., e_{n-1} g^{-1}\}$ and can be visualized as a reflection across this subspace.
(3) Prove that every element of $O(2) - SO(2)$ is a reflection, and every element of $SO(2)$ is the product of two reflections.
(4) Prove that every element of the standard maximal torus of $SO(n)$ is the product of finitely many reflections.
(5) Prove that every element of $O(n)$ is the product of finitely many reflections.

**Ex. 9.7.** Identify $Sp(1)$ with the unit-length quaternions $S^3 \subset \mathbb{H}$.

(1) Prove that the conjugates of the standard maximal torus of $Sp(1)$ are exactly the intersections of $S^3$ with the 2-dimensional $\mathbb{R}$-subspaces of $\mathbb{H}$ that contain 1.

(2) Prove that two elements $a_1+b_1\mathbf{i}+c_1\mathbf{j}+d_1\mathbf{k}, a_2+b_2\mathbf{i}+c_2\mathbf{j}+d_2\mathbf{k}$ in $Sp(1)$ are conjugate if and only if $a_1 = a_2$.

(3) Prove that two elements of $SU(2)$ are conjugate if and only if they have the same trace.
*Hint: Consider the isomorphism* $\Psi_1 : Sp(1) \to SU(2)$.

**Ex. 9.8.** If $H \subset G$ is a subgroup, define the **normalizer** of $H$ as $N(H) = \{g \in G \mid gHg^{-1} = H\}$. Prove that $N(H)$ is a subgroup of $G$ and that $H$ is a normal subgroup of $N(H)$.

**Ex. 9.9.** Let $G \in \{SO(n), U(n), SU(n), Sp(n)\}$, let $T$ be the standard maximal torus of $G$, and let $\tau \subset \mathfrak{g}$ denote their Lie algebras.

(1) Prove that if $X \in \mathfrak{g}$ commutes with every vector in $\tau$, then $X \in \tau$. In other words, $\tau$ is a "maximal abelian" subspace.

(2) Prove that the Lie algebra of $N(T)$ equals $\tau$.
*Hint: Use part (1) and also Exercise* 8.13.

(3) Conclude that $N(T)$ is comprised of finitely many nonintersecting subsets of $G$, each diffeomorphic to $T$.

**Ex. 9.10.** Prove that the normalizer of the standard maximal torus $T$ of $Sp(1)$ is:

$$N(T) = T \cup (T \cdot \mathbf{j}).$$

**Ex. 9.11.** Prove that the normalizer of the standard maximal torus $T$ of $SO(3)$ is:

$$N(T) = T \cup \left\{ \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ \sin\theta & -\cos\theta & 0 \\ 0 & 0 & -1 \end{pmatrix} \middle| \theta \in [0, 2\pi) \right\}.$$

Chapter 10

# Homogeneous manifolds

Other than spheres and matrix groups, we haven't yet encountered many examples of manifolds. In this chapter, we will construct many more, including the projective spaces, which are of fundamental importance in many areas of physics and mathematics. Our goal is to describe new manifolds in simple and elegant ways rather than with messy equations. The only way to achieve this goal is to first generalize our definition of "manifold."

## 1. Generalized manifolds

In Chapter 7, we defined an *embedded manifold* to essentially mean a subset of some ambient Euclidean space that is locally identified with $\mathbb{R}^d$. The name indicates that it is "embedded" in the ambient Euclidean space. We will henceforth use the word "manifold" to mean a slightly more general type of object. We will define a (generalized) manifold to essentially mean a *set* that is locally identified with $\mathbb{R}^d$. There will no longer be an ambient Euclidean space. This generalization is motivated by several considerations:

(1) Cosmologists model the universe as a 3-dimensional manifold, but it would be unnatural for the model to be a subset of $\mathbb{R}^4$ or $\mathbb{R}^5$, since the universe is the whole universe.

(2) In Section 6 of Chapter 8, we defined $\mathbb{RP}^n$ as the set of all lines through the origin in $\mathbb{R}^{n+1}$, and we described a sense in which $\mathbb{RP}^n$ locally looks like $\mathbb{R}^n$, and therefore deserves to be called a manifold, even though it is a set of lines (rather than a set of points in some Euclidean space).

(3) In Section 1 of Chapter 9, we proved that

$$(\mathbb{R}^n, +)/\langle\{e_1, ..., e_n\}\rangle$$

is isomorphic to the torus, $T^n$. In fact, it's *smoothly* isomorphic, but for this assertion to make sense, we must somehow regard this coset space as a manifold. More generally, many important manifolds are best described as coset spaces of the form $G/H$, where $G$ is a matrix group and $H \subset G$ is a closed subgroup. We must learn to regard a coset space as a manifold, even though it is a set of cosets (rather than a set of points in Euclidean space).

**Definition 10.1.** *A* ***manifold*** *of dimension $d$ is set, $M$, together with a family of injective functions $\varphi_i : U_i \to M$, called* ***parametrizations*** *(where each $U_i$ is an open set in $\mathbb{R}^d$) such that:*

(1) $\bigcup_i \varphi_i(U_i) = M$ *(the parametrizations cover all of $M$).*

(2) (Compatibility condition) *For any pair $i, j$ of indices with $W = \varphi_i(U_i) \cap \varphi_j(U_j) \neq \emptyset$, the sets $\varphi_i^{-1}(W)$ and $\varphi_j^{-1}(W)$ are open in $\mathbb{R}^d$, and*

$$\varphi_j^{-1} \circ \varphi_i : \varphi_i^{-1}(W) \to \varphi_j^{-1}(W)$$

*is a smooth function between these open sets.*

(3) *The family of parametrizations is maximal with respect to conditions (1) and (2).*

*The family of parametrizations is called an* ***atlas*** *for $M$.*

The compatibility condition is illustrated in Figure 1. This condition is satisfied for the embedded manifolds defined in Chapter 7 (basically because $\varphi_j^{-1} \circ \varphi_i$ is a composition of two smooth functions, but the reader will check the details in Exercise 10.1).

Condition (3) insures, for example, that the restriction of any parametrization to any open subset of its domain is also a parametri-

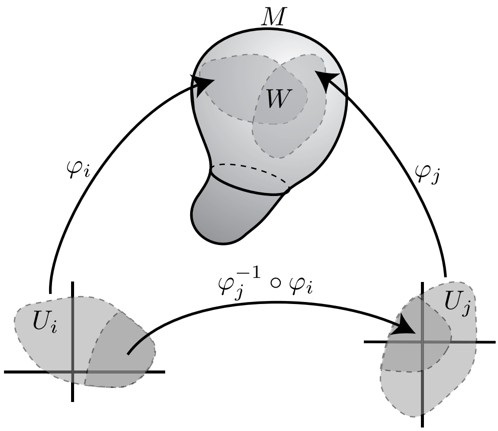



**Figure 1.** The change of parametrization, $\varphi_j^{-1} \circ \varphi_i$, must be smooth.

zation of the atlas. Our focus in this chapter is the construction of new examples of manifolds. For this activity, we can safely ignore condition (3); it will be enough to construct an atlas satisfying (1) and (2). This is because such an atlas can be uniquely completed to an atlas also satisfying (3). The completion is achieved by adding to it all parametrizations compatible with the original ones in the sense of (2). These newly added parametrizations will automatically also be compatible with each other (Exercise 10.2).

Since a manifold is locally identified (via parametrizations) with open sets in $\mathbb{R}^d$, one can do to a manifold most things that can be done to an open subset of Euclidean space. We'll begin by discussing how to do topological things:

**Definition 10.2.** *Let $M$ be a manifold. A subset of $M$ is called **open** (in $M$) if it equals a union of sets of the form $\varphi_i(A)$, where $A \subset U_i$ is an open subset of the domain, $U_i$, of a parametrization $\varphi_i : U_i \to M$. The collection of all open sets in $M$ is called the **topology** of $M$.*

In particular, the image of any parametrization covering a point of $M$ is a **neighborhood** of that point (which means an open set in $M$ containing that point).

In Chapter 3, we observed that the following concepts could be defined purely in terms of topology: convergence, limit points, closed, compact, path-connected, and continuous. Therefore all of these definitions (previously stated for subsets of Euclidean space) make perfect sense for manifolds. For example, a function $f : M_1 \to M_2$ between manifolds is called *continuous* if for every set $A$ that's open in $M_2$, $f^{-1}(A)$ is open in $M_1$.

We next use local parameterizations to define, among other things, the *derivative* of a function between manifolds:

**Definition 10.3.** *Let $M_1$ and $M_2$ be manifolds with dimensions $d_1$ and $d_2$ respectively. Let $f : M_1 \to M_2$ be a continuous function, and let $p \in M_1$. Choose parametrizations $\varphi_1 : U_1 \to V_1 \subset M_1$ and $\varphi_2 : U_2 \to V_2 \subset M_2$ covering $p$ and $f(p)$ respectively. Assume that $f(V_1) \subset V_2$; if this is not already the case, it can be achieved by replacing $U_1$ with the smaller open set $\varphi_1^{-1}(V_1 \cap f^{-1}(V_2))$. Define $\phi = \varphi_2^{-1} \circ f \circ \varphi_1 : U_1 \to U_2$. Define $x = \varphi_1^{-1}(p) \in U_1$.*

(1) *$f$ is called **differentiable** at $p$ if $\phi$ is differentiable at $x$.*

(2) *A "**curve** in $M_1$ through $p$" is a function $\gamma : (-\epsilon, \epsilon) \to M_1$ (for some $\epsilon > 0$) such that $\gamma(0) = p$ and $\gamma$ is differentiable at $0$ (in the sense of part (2)).*

(3) *A "**tangent vector** to $M_1$ at $p$" is an equivalence class of curves in $M_1$ through $p$, with two curves considered equivalent if their compositions with $\varphi_1^{-1}$ (which are curves in $U_1$ through $x$) have the same initial derivative.*

(4) *The "**tangent space** to $M_1$ at $p$," denoted $T_pM_1$, is the set of all tangent vectors to $M_1$ at $p$, endowed with the structure of a vector space (addition and scalar multiplication) via the identification:*

$$T_pM_1 \cong T_xU_1 \cong \mathbb{R}^{d_1} \qquad \text{described as:}$$
$$[\gamma] \leftrightarrow (\varphi_1^{-1} \circ \gamma)'(0),$$

*where $[\gamma] \in T_pM_1$ denotes the equivalence class of the curve $\gamma$ in $M_1$ through $p$.*

(5) *(Assuming $f$ is differentiable at $p$) the **derivative** of $f$ at $p$ is the linear function*

$$df_p : T_pM_1 \to T_{f(p)}M_2$$

*defined as*

$$df_p([\gamma]) = [f \circ \gamma].$$

*Identifying $T_pM_1 \cong \mathbb{R}^{d_1}$ and $T_{f(p)}M_2 \cong \mathbb{R}^{d_2}$ as in part (5), $df_p$ can equivalently be defined as the linear function identified with $d\phi_x : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$.*

In Figure 2, an element $[\gamma] \in T_pM_1$ is illustrated as a vector in space, which is literally correct only for an embedded manifold in $\mathbb{R}^3$, but is nevertheless a useful way to picture tangent vectors generally.

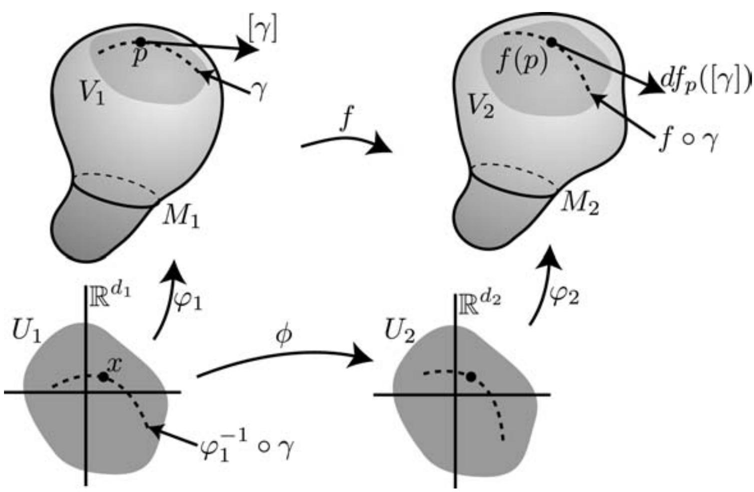



**Figure 2.** $df_p([\gamma]) = [f \circ \gamma]$.

We leave it to the reader to confirm that Definition 10.3 is well-defined; that is, replacing $\varphi_1$ and $\varphi_2$ with different parametrizations covering $p$ and $f(p)$ respectively would NOT change:

- whether $f$ is continuous (or differentiable) at $p$.
- which pairs of curves in $M_1$ through $p$ are equivalent.

- the vector space operations (addition and scalar multiplication) on $T_pM_1$.

The compatibility condition from Definition 10.1 is contrived to insure all of these things.

The reader should also check that part (5) is well defined in the sense that the element $[f \circ \gamma] \in T_{f(p)}M_2$ does not depend on the representative, $\gamma$, of the equivalence class $[\gamma] \in T_pM_1$.

Definition 10.3 reduces to the previous definitions of these terms from Chapter 7 in the special case of embedded manifolds. For example, when $M_1 \subset \mathbb{R}^m$ is an embedded manifold, two curves in $M_1$ through $p$ are equivalent if and only if their initial velocities equal the same element of $\mathbb{R}^m$. Therefore, elements of the newly defined $T_pM_1$ naturally correspond one-to-one with elements of the previously defined $T_pM_1$ via the correspondence $[\gamma] \leftrightarrow \gamma'(0)$.

A function $f : M_1 \to M_2$ between two manifolds is called **smooth** if for every $p \in M_1$, there exist parametrizations covering $p$ and $f(p)$ such that $\phi$ (defined as in Definition 10.3) is smooth. For embedded manifolds, it requires some work to prove that this is equivalent to our previous definition of smoothness. Exactly as before, two manifolds are called **diffeomorphic** if there exists a smooth bijection between them whose inverse is also smooth.

The same strategy can be used to decide whether a vector field is smooth. A **vector field**, $X$, on a manifold, $M$, means a choice for each $p \in M$ of a vector $X(p) \in T_pM$. Given a parametrization, $\varphi : U \subset \mathbb{R}^d \to V \subset M$, the restriction of $X$ to $V$ is naturally associated with a function from $U$ to $\mathbb{R}^d$. The vector field $X$ is called **smooth** if each point of $M$ is covered by a parametrization with respect to which this associated function is smooth.

The **Whitney Embedding Theorem** says that any manifold is diffeomorphic to an embedded manifold. This means that the class of generalized manifolds is really no more general than the class of embedded manifolds that we previously studied in Chapter 7. However, the specific manifolds we will study throughout this chapter would not look elegant or simple if one tried to describe them as embedded manifolds. Describing them as generalized manifolds is the way to go.

## 2. The projective spaces

Other than spheres, the most important manifolds in geometry are the projective spaces, which we will define and study in this section.

**Definition 10.4.** *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and $n \geq 1$. A* ***line*** *in $\mathbb{K}^{n+1}$ means a 1-dimensional $\mathbb{K}$-subspace of $\mathbb{K}^{n+1}$, that is, a set that has the form $\{\lambda \cdot v \mid \lambda \in \mathbb{K}\}$ for some non-zero vector $v \in \mathbb{K}^{n+1}$. The set*

$$\mathbb{KP}^n = \{\textit{all lines in } \mathbb{K}^{n+1}\}$$

*is called* ***real projective space*** $(\mathbb{RP}^n)$*,* ***complex projective space*** $(\mathbb{CP}^n)$*, or* ***quaternionic projective space*** $(\mathbb{HP}^n)$*.*

Under the identification $\mathbb{C}^{n+1} \cong \mathbb{R}^{2(n+1)}$, notice that any line in $\mathbb{C}^{n+1}$ gets identified with a 2-dimensional $\mathbb{R}$-subspace of $\mathbb{R}^{2(n+1)}$ and is therefore best visualized as a plane rather than a line. Similarly, a line in $\mathbb{H}^{n+1}$ gets identified with a 4-dimensional $\mathbb{R}$-subspace of $\mathbb{R}^{4(n+1)}$.

**Example 10.5** (A manifold structure for $\mathbb{KP}^n$)**.** *For any given point $(z_0, z_1, ..., z_n) \in \mathbb{K}^{n+1}$ (other than the origin), the unique line containing this point will be denoted as $[z_0, z_1, ..., z_n] \in \mathbb{KP}^n$.*

*For each $i = 0, 1, ..., n$, define $V_i = \{[z_0, z_1, ..., z_n] \in \mathbb{KP}^n \mid z_i \neq 0\}$ and define $\varphi_i : \mathbb{K}^n \to V_i$ to be the map that assigns the $i^{th}$ coordinate to equal 1; that is:*

$$\begin{aligned}
\varphi_0(z_1, z_2, ..., z_n) &= [1, z_1, z_2, ..., z_n], \\
\varphi_1(z_1, z_2, ..., z_n) &= [z_1, 1, z_2, ..., z_n], \\
&\ \ \vdots \\
\varphi_n(z_1, z_2, ..., z_n) &= [z_1, z_2, ..., z_n, 1].
\end{aligned}$$

*These $n+1$ parametrizations are injective and together they cover all of $\mathbb{KP}^n$.*

*It remains to verify the compatibility condition of Definition* 10.1. *For example, when $i = 0$ and $j = 1$,*

$$W = \varphi_0(\mathbb{K}^n) \cap \varphi_1(\mathbb{K}^n) = \{[z_0, z_1, ..., z_n] \in \mathbb{KP}^n \mid z_0, z_1 \neq 0\},$$

*and therefore*

$$\varphi_0^{-1}(W) = \varphi_1^{-1}(W) = \{(z_1, ..., z_n) \in \mathbb{K}^n \mid z_1 \neq 0\},$$

*which is open. Furthermore,*

$$\left(\varphi_1^{-1} \circ \varphi_0\right)(z_1, z_2, ..., z_n) = (z_1^{-1}, z_2, ..., z_n),$$

*which is smooth. There is nothing special here about $i = 0, j = 1$; condition (2) can be similarly verified for all other pairs of indices.*

*Since these parametrizations locally identify $\mathbb{KP}^n$ with $\mathbb{K}^n$, we see that dim$(\mathbb{KP}^n) \in \{n, 2n, 4n\}$, depending on $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$.*

There is a noteworthy linguistic difference between an embedded manifold (Definition 7.14) and a general manifold (Definition 10.1). One may ask whether a subset of Euclidean space *is* an embedded manifold, whereas one may only ask whether an arbitrary set "has a natural manifold structure," meaning an atlas as in the definition. In the previous example, we didn't prove that $\mathbb{KP}^n$ *is* a manifold, but rather we made it into a manifold in a natural way. The symbol "$\mathbb{KP}^n$" will henceforth denote the set of lines in $\mathbb{K}^{n+1}$ together with the above-described manifold structure (atlas).

For $n \geq 2$, the spaces $\mathbb{CP}^n$ and $\mathbb{HP}^n$ are not diffeomorphic to any previously familiar manifolds, but $\mathbb{CP}^1$ and $\mathbb{HP}^1$ are very familiar:

**Proposition 10.6.** $\mathbb{CP}^1$ *is diffeomorphic to $S^2$, and* $\mathbb{HP}^1$ *is diffeomorphic to $S^4$.*

**Proof.** We will prove the first claim; the second is similar and is left to the reader in Exercise 10.9. We will show that the bijection $F : S^2 \to \mathbb{CP}^1$ defined so that for all unit-length vectors $(x, y, z)$,

$$F(x, y, z) = \begin{cases} [1, 0] & \text{if } z = 1 \\ \left[\frac{2}{1-z}(x + \mathbf{i}y), 1\right] & \text{if } z \neq 1 \end{cases}$$

is a diffeomorphism. For inputs other than $(0,0,1)$, $F$ is the composition of the following three diffeomorphisms:

$$\begin{array}{ccc} S^2 - \{(0,0,1)\} & & \mathbb{CP}^1 - \{[1,0]\} \\ {\scriptstyle (x,y,z)\mapsto\frac{2}{1-z}(x,y)}\Big\downarrow & & \Big\uparrow{\scriptstyle (a+b\mathbf{i})\mapsto[a+b\mathbf{i},1]} \\ \mathbb{R}^2 & \xrightarrow{(a,b)\mapsto(a+\mathbf{i}b)} & \mathbb{C} \end{array}$$

The left function is stereographic projection, defined in Exercise 7.1. For these inputs, the definition of smoothness is obviously satisfied with respect to the parametrizations determined by the vertical arrows in the above diagram.

It remains to choose parametrizations covering the exceptional point $p = (0,0,1)$ and its image $f(p) = [1,0]$, and to verify with respect to these parametrizations that $F$ and its inverse satisfy the definition of smooth at $p$. For this, we choose the parametrizations determined by the vertical arrows in this diagram:

$$\begin{array}{ccc} S^2 - \{(0,0,-1)\} & \xrightarrow{F} & \mathbb{CP}^1 - \{[0,1]\} \\ {\scriptstyle (x,y,z)\mapsto\frac{2}{1+z}(x,y)}\Big\downarrow & & \Big\uparrow{\scriptstyle (a+b\mathbf{i})\mapsto[1,a+b\mathbf{i}]} \\ \mathbb{R}^2 & \xrightarrow{\phi} & \mathbb{C} \end{array}$$

The left function is the projection onto the plane $z = 1$ discussed in Exercise 7.1(3). The reader should check that $\phi : \mathbb{R}^2 \to \mathbb{C}$ (defined as the composition: up, then right, then down) is described by the formula

$$\phi(x,y) = \frac{1}{4}(x - \mathbf{i}y),$$

which is clearly smooth with smooth inverse. □

The next section is devoted to proving the following: if $G$ is a matrix group and $H \subset G$ is a closed subgroup, then the coset space $G/H$ has a natural manifold structure. As a clarifying illustration, we end the current section with a description of the natural manifold structure on one particular coset space:

**Example 10.7** ($Sp(1)/T \cong S^2$). *Regard $Sp(1) \cong S^3 \subset \mathbb{H}$ as the group of unit-length quaternions. Its standard maximal torus is the circle:*

$$T = \{e^{\mathbf{i}\theta} \mid \theta \in [0,2\pi)\} = S^3 \cap \mathbb{C},$$

*where* $\mathbb{C} \subset \mathbb{H}$. *For any* $q \in S^3$, *the coset* $q \cdot T \in S^3/T$ *is:*

$$q \cdot T = \{q \cdot e^{\mathbf{i}\theta} \mid \theta \in [0, 2\pi)\} = S^3 \cap \{q \cdot \lambda \mid \lambda \in \mathbb{C}\}, \tag{10.1}$$

*which is also a circle; in fact, it is diffeomorphic to* $T$ *via the diffeomorphism that left-multiplies by* $q$. *For example,* $\mathbf{j} \cdot T$ *is the circle in the* $\{\mathbf{j}, \mathbf{k}\}$*-plane of* $\mathbb{H}$. *In general, the cosets of any subgroup partition the group; in this case, it is interesting that the 3-dimensional sphere is naturally partitioned into a collection of disjoint circles.*

*These are not just any circles. It is apparent from Equation* 10.1 *that each coset is a "complex great circle," which means the intersection of* $S^3$ *with a 1-dimensional* $\mathbb{C}$*-subspace of* $\mathbb{H} \cong \mathbb{C}^2$ *(see Exercise* 2.12 (2) *for a description of the identification* $\mathbb{H} \cong \mathbb{C}^2$ *that is appropriate here). The coset space* $S^3/T$ *is thereby identified with the set of lines in* $\mathbb{C}^2$, *which is called* $\mathbb{CP}^1$, *and which by Proposition* 10.6 *is diffeomorphic to* $S^2$.

## 3. Coset spaces are manifolds

When $G$ is a group and $H \subset G$ is a subgroup, recall that "$G/H$" and "$H\backslash G$" respectively denote the left and right *coset space*, i.e., the collection of (left/right) cosets. The left coset containing $g$ will be denoted either as $g \cdot H$ or as $[g]$. The right coset containing $g$ will be denoted either as $H \cdot g$ or as $[g]$. Algebra textbooks focus on the case where $H$ is normal in $G$, so that the coset space is itself a group. We will not care whether $H$ is normal in $G$, but we will instead assume that $G$ and $H$ are matrix groups; in this case, the coset space has a natural manifold structure:

**Theorem 10.8.** *Let* $G$ *be a matrix group and* $H \subset G$ *a closed subgroup. Let* $\mathfrak{h} \subset \mathfrak{g}$ *denote their Lie algebras. Define*

$$\mathfrak{p} = \mathfrak{h}^\perp = \{V \in \mathfrak{g} \mid \langle V, X \rangle_{\mathbb{R}} = 0 \ \ \forall X \in \mathfrak{h}\} \ \ \textit{and} \ \ \mathfrak{p}_\epsilon = \{V \in \mathfrak{p} \mid |V| < \epsilon\}.$$

*For every* $g \in G$, *define the parametrization* $\varphi_g : \mathfrak{p}_\epsilon \to G/H$ *as:*

$$\varphi_g(X) = [g \cdot \exp(X)].$$

*If* $\epsilon$ *is sufficiently small, the family of parametrizations* $\{\varphi_g \mid g \in G\}$ *determines a manifold structure on the coset space* $G/H$; *that is, they are injective and they satisfy the compatibility condition of Definition* 10.1.

The notation "$G/H$" will henceforth denote the left coset space together with the manifold structure described in the theorem. Notice that the dimension of the $G/H$ equals

$$\dim(G/H) = \dim(\mathfrak{p}) = \dim(G) - \dim(H).$$

The theorem and and its proof can be modified in the obvious way to provide a manifold structure on the *right* coset space, with parametrizations defined as $\varphi_g(X) = [\exp(X) \cdot g]$. The notation "$H\backslash G$" will henceforth denote the right coset space together with this manifold structure.

Most of the work of the proof goes into the following lemma, which is of independent interest:

**Lemma 10.9.** *With the assumptions and notation of Theorem* 10.8*, there exists* $\epsilon > 0$ *such that the function* $\Psi : \mathfrak{p}_\epsilon \times H \to G$ *defined as*

$$\Psi(X, h) = \exp(X) \cdot h$$

*is a diffeomorphism onto its image, which is an open set in* $G$ *containing* $H$*.*

**Proof of Lemma 10.9.** The derivative of $\Psi$ at $(0, I)$ is the identity, so the Inverse Function Theorem (Theorem 7.22) implies that the restriction of $\Psi$ to a sufficiently small neighborhood of $(0, I)$ is a diffeomorphism onto its image. There exists $\epsilon > 0$ and a neighborhood, $U$, of $I$ in $H$ such that the restriction of $\Psi$ to $\mathfrak{p}_\epsilon \times U$ is a diffeomorphism onto its image (because $\epsilon$ and $U$ can be chosen such that this domain lies inside the previously mentioned neighborhood). For any $h \in H$, the right-multiplication map $\mathcal{R}_h : G \to G$ is a diffeomorphism, so the restriction of $\Psi$ to $\mathfrak{p}_\epsilon \times (U \cdot h)$ is also a diffeomorphism onto its image.

In summary, we have chosen a value $\epsilon > 0$ such that $\Psi : \mathfrak{p}_\epsilon \times H \to G$ is a *local* diffeomorphism at each point of its domain (which means its restriction to a neighborhood of that point is a diffeomorphism onto its image).

To prove the theorem, it remains to show that $\Psi$ is injective for some possibly smaller choice of $\epsilon$. Suppose to the contrary that no such $\epsilon$ exists; that is, for every $\epsilon > 0$, there exists a pair of *distinct* points $(X_1, h_1), (X_2, h_2) \in \mathfrak{p}_\epsilon \times H$ such that

$$\begin{aligned}\Psi(X_1, h_1) = \Psi(X_2, h_2) &\iff \exp(X_1) \cdot h_1 = \exp(X_2) \cdot h_2 \\ &\iff \exp(X_2)^{-1} \cdot \exp(X_1) = h_2 h_1^{-1} \in H.\end{aligned}$$

Theorem 7.1 applied to $H$ (which is itself a matrix group because it is closed) implies there exists $A \in \mathfrak{h}$ with $\exp(A) = h_2 h_1^{-1}$, provided $\epsilon$ is sufficiently small. The norm of $A$ can be made arbitrarily small by choosing $\epsilon$ arbitrarily small. In summary:

$$\exp(X_2)^{-1} \cdot \exp(X_1) = \exp(A). \tag{10.2}$$

Now consider the smooth function $F : \mathfrak{g} \to G$ defined as

$$F(Y) = \exp(Y^{\mathfrak{p}}) \cdot \exp(Y^{\mathfrak{h}}),$$

where the superscripts denote orthogonal projections onto those subspaces. The derivative of $F$ at 0 is the identity map, so according to the Inverse Function Theorem (Theorem 7.22), $F$ is a diffeomorphism when restricted to a sufficiently small neighborhood of the origin in $\mathfrak{g}$. Now re-write Equation 10.2 as:

$$\underbrace{\exp(X_1)}_{F(X_1+0)} = \underbrace{\exp(X_2) \cdot \exp(A)}_{F(X_2+A)}.$$

When $\epsilon$ is sufficiently small, this contradicts the injectivity of $F$. $\square$

**Proof of Theorem 10.8.** Choose $\epsilon$ as in Lemma 10.9. For any $g \in G$, $\varphi_g$ is injective because for any $X_1, X_2 \in \mathfrak{p}_\epsilon$:

$$\begin{aligned}\varphi_g(X_1) = \varphi_g(X_2) &\iff [g \cdot \exp(X_1)] = [g \cdot \exp(X_2)] \\ &\iff \exp(X_1)^{-1} \cdot \exp(X_2) = h \in H \\ &\iff \exp(X_2) = \exp(X_1) \cdot h \\ &\iff \Psi(X_2, I) = \Psi(X_1, h) \\ &\iff X_1 = X_2 \text{ and } h = I.\end{aligned}$$

It remains to verify the compatibility condition of Definition 10.1. Let $a, b \in G$ such that $W = \varphi_a(\mathfrak{p}_\epsilon) \cap \varphi_b(\mathfrak{p}_\epsilon) \neq \emptyset$. For $X \in \mathfrak{p}_\epsilon$, notice that $X \in \varphi_a^{-1}(W)$ if and only if there exists $(Y, h) \in \mathfrak{p}_\epsilon \times H$ such that:

$$a \cdot \exp(X) = b \cdot \exp(Y) \cdot h \iff \exp(X) = (a^{-1}b) \cdot \Psi(Y, h). \tag{10.3}$$

So $\varphi_a^{-1}(W)$ is open because it equals the pre-image under the continuous function $\exp : \mathfrak{p}_\epsilon \to G$ of the open set $(a^{-1}b) \cdot \text{Image}(\Psi)$.

Lemma 10.9 implies that the choice $(Y, h) \in \mathfrak{p}_\epsilon \times H$ in Equation 10.3 is uniquely and smoothly determined by $X$. The transition function $\varphi_b^{-1} \circ \varphi_a : \varphi_a^{-1}(W) \to \varphi_b^{-1}(W)$, which sends $X \mapsto Y$, is therefore smooth. □

**Proposition 10.10.** *With the assumptions and notation of Theorem* 10.8*, define* $\pi : G \to G/H$ *such that* $\pi(g) = [g]$ *for all* $g \in G$*.*

(1) $\pi$ *is smooth.*
(2) *When* $M$ *is a manifold, a function* $f : G/H \to M$ *is smooth if and only if* $f \circ \pi$ *is smooth.*

**Proof.** Exercise 10.10. □

## 4. Group actions

To understand the symmetries of a manifold, one must understand how groups act on it.

**Definition 10.11.** *A (left or right)* ***action*** *of a group* $G$ *on a set* $M$ *is a function*

$$\varphi : G \to \{\textit{the set of bijections from } M \textit{ to } M\}$$

*such that for all* $g_1, g_2 \in G$*,*

$$\begin{aligned} &\textit{(left action)} \quad \varphi(g_1 \cdot g_2) = \varphi(g_1) \circ \varphi(g_2), \\ &\textit{(right action)} \quad \varphi(g_1 \cdot g_2) = \varphi(g_2) \circ \varphi(g_1). \end{aligned}$$

*An action of a group* $G$ *on a vector space over* $\mathbb{K}$ *is called a* ***$\mathbb{K}$-linear*** *action (or a* ***representation****) if each bijection in its image is* $\mathbb{K}$*-linear.*

*An action of a matrix group* $G$ *on a manifold* $M$ *is called* ***smooth*** *if the function* $G \times M \to M$ *that sends* $(g, p) \mapsto \varphi(g)(p)$ *is smooth (with respect to the natural manifold structure on* $G \times M$ *discussed in Exercise* 10.5*). This implies that each bijection in its image is a diffeomorphism.*

The set of bijections from $M$ to $M$ form a group under composition. A *left* action of $G$ on $M$ could equivalently be defined as a homomorphism from $G$ to this group.

We have already encountered and studied the following important group actions:

**Example 10.12.** *Let $G \subset GL_n(\mathbb{K})$ be a matrix group with Lie algebra denoted $\mathfrak{g}$.*

(1) *The action of $G$ on $\mathbb{K}^n$ by left multiplication ($\varphi(g) = L_g$) is a left action. It is $\mathbb{R}$-linear but not necessarily $\mathbb{K}$-linear.*

(2) *The action of $G$ on $\mathbb{K}^n$ by right multiplication ($\varphi(g) = R_g$) is a $\mathbb{K}$-linear right action.*

(3) *If $G \subset \mathcal{O}_n(\mathbb{K})$, then the previous two actions of $G$ on $\mathbb{K}^n$ restrict to smooth actions on the sphere of unit-length vectors in $\mathbb{K}^n$.*

(4) *The adjoint action of $G$ on $\mathfrak{g}$ ($\varphi(g) = Ad_g$) is a left $\mathbb{R}$-linear action. If $G \subset \mathcal{O}_n(\mathbb{K})$, then it restricts to a smooth left action on the sphere of unit-length vectors in $\mathfrak{g}$.*

(5) *There are three natural smooth actions of $G$ on $G$:*
    (a) *Left-multiplication ($\varphi(g) = \mathcal{L}_g$) is a smooth left action.*
    (b) *Right-multiplication ($\varphi(g) = \mathcal{R}_g$)is a smooth right action.*
    (c) *Conjugation ($\varphi(g) = C_g$) is a smooth left action.*

Group actions are fundamentally important to many fields of mathematics. The following general vocabulary will be useful as we explore their relevance to matrix groups:

**Definition 10.13.** *Let $\varphi$ be a (left or right) action of a group $G$ on a set $M$.*

(1) *For any $g \in G$ and $p \in M$, the element $\varphi(g)(p) \in M$ will also be denoted as "$g \star p$".*

(2) *For any $p \in M$, the set $G \star p = \{g \star p \mid g \in G\}$ is called the* ***orbit*** *containing $p$.*

(3) *The set of orbits is called the* ***orbit space*** *and is denoted as $G \backslash M$ (for a left action) or $M/G$ (for a right action).*

(4) *$\varphi$ is called **transitive** if there is only one orbit; that is, for any pair $p, q \in M$ there exists $g \in G$ such that $g \star p = q$.*

(5) *For any $p \in G$, the set $G_p = \{g \in G \mid g \star p = p\}$ is called the **stabilizer** of $p$.*

(6) *$\varphi$ is called **free** if all stabilizers are trivial; in other words, $(g \star p = p) \Rightarrow (g = I)$.*

For a smooth action of a matrix group $G$ on a manifold $M$, all stabilizers are *closed* subgroups of $G$ (Exercise 10.14).

**Example 10.14.** *Let $G$ be a matrix group and $H \subset G$ a closed subgroup. The action of $H$ on $G$ by right multiplication is a smooth free right action whose orbits are the left cosets, so the orbit space equals the coset space. The familiar notation "$G/H$" for this coset space is consistent with our new notation for a general orbit space. Furthermore, the action of $G$ on the orbit space $G/H$ defined as*

$$g_1 \star [g_2] = [g_1 \cdot g_2]$$

*is a well-defined transitive left action. Proposition* 10.10 *can be used to prove that this action is smooth (Exercise* 10.16*).*

*Similarly, the action of $H$ on $G$ by left multiplication is a smooth free left action whose orbit space, $H\backslash G$, is the space of right cosets. The action of $G$ on this coset space defined as $g_1 \star [g_2] = [g_2 \cdot g_1]$ is a well-defined transitive smooth right action.*

In summary: If $G$ is a matrix group and $H \subset G$ a closed subgroup, then either coset space ($G/H$ or $H\backslash G$) is a special type of manifold – one on which a matrix group acts transitively.

**Definition 10.15.** *A manifold, $M$, is called **homogeneous** if there exists a transitive smooth action of a matrix group on $M$.*

Conversely, it turns out essentially that every homogeneous manifold is diffeomorphic to a coset space. We will sketch the proof of this assertion in the next section.

## 5. Homogeneous manifolds

In this section, as promised, we will sketch a proof that every homogeneous manifold is diffeomorphic to a coset space. To illustrate the key idea, it is useful to first pursue a different goal: to describe an unexpected way in which Theorem 10.8 can provide a homogeneous manifold structure on certain sets. The first set that we'd like to turn into a homogeneous manifold is:

**Definition 10.16.** *Let $m < n$ be positive integers. The set of $m$-dimensional $\mathbb{K}$-subspaces of $\mathbb{K}^n$ is denoted $G_m(\mathbb{K}^n)$ and is called the* ***Grassmann manifold*** *of $m$-planes in $\mathbb{K}^n$.*

The name suggests that it's a manifold, but it is not obvious how to construct an atlas. Luckily we won't have to. Instead, consider the single element $V_0 = \text{span}\{e_1, e_2, ..., e_m\} \in G_m(\mathbb{K}^n)$ (the span of the first $m$ members of the standard orthonormal basis of $\mathbb{K}^n$). For any $g \in GL_n(\mathbb{K})$, notice that $R_g(V_0) = \{v \cdot g \mid v \in V_0\}$ is another element of $G_m(\mathbb{K}^n)$. But this observation doesn't quite allow us to identify $GL_n(\mathbb{K})$ with $G_m(\mathbb{K}^n)$ because the identification is not one-to-one. For example, $R_g(V_0) = V_0$ if and only if $g$ lies in the following subgroup:

$$H_0 = \{g \in GL_n(\mathbb{K}) \mid \text{each of the first } m \text{ rows of } g \text{ is in } V_0\}. \tag{10.4}$$

That is, $H_0$ contains all the general linear matrices whose top-right block equals the $m$-by-$(n-m)$ zero matrix.

In fact, $R_{g_1}(V_0) = R_{g_2}(V_0)$ if and only if $g_1$ and $g_2$ lie in the same coset of $H_0 \backslash GL_n(\mathbb{K})$. We therefore have a natural identification

$$G_m(\mathbb{K}^n) \cong H_0 \backslash GL_n(\mathbb{K}).$$

Theorem 10.8 provides a homogeneous manifold structure on this coset space and (via this identification) also on $G_m(\mathbb{K}^n)$.

The key idea here is the following simple and general fact:

**Lemma 10.17.** *Let $\varphi$ be a transitive left (respectively right) action of a group $G$ on a set $M$, let $p_0 \in M$, and let $H = G_{p_0}$ be the stabilizer of $p_0$. Then the function $F : G/H \to M$ (respectively $F : H\backslash G \to M$) defined as*

$$F([g]) = g \star p_0$$

*is a well-defined bijection.*

**Proof.** Exercise 10.15. □

In either case, $F$ is in fact an **equivariant** bijection, which means that it identifies the transitive actions of $G$ on these two sets. This means, for example in the left version, that for each $g \in G$, the following diagram commutes:

$$\begin{tikzcd} G/H \arrow[r, "\tilde{\varphi}(g)"] \arrow[d, "F"'] & G/H \arrow[d, "F"] \\ M \arrow[r, "\varphi(g)"] & M \end{tikzcd} \tag{10.5}$$

where $\tilde{\varphi}$ is the action of $G$ on $G/H$ defined as $\tilde{\varphi}(g)([x]) = [g \cdot x]$, while $\varphi$ is the given action of $G$ on $M$, denoted as $\varphi(g)(p) = g \star p$.

According to Exercise 10.12, no generality is lost in developing the general theory only for *left* actions. Nevertheless, certain examples are most naturally described in terms of right actions, including the next example.

**Example 10.18** (Steifel manifolds)**.** *Let $m < n$ be positive integers. The **Steifel manifold** of $m$-frames in $\mathbb{K}^n$, denoted $S_m(\mathbb{K}^n)$, is defined as the set of all ordered lists of $m$ linearly independent elements of $\mathbb{K}^n$.*

*There is a natural transitive right action of $GL_n(\mathbb{K})$ on $S_m(\mathbb{K}^n)$ described as follows: if $g \in GL_n(\mathbb{K})$ and $F = (v_1, ..., v_m) \in S_m(\mathbb{K}^n)$, then $g \star F = (v_1 \cdot g, ..., v_m \cdot g) \in S_m(\mathbb{K}^n)$. The simplest element of $S_m(\mathbb{K}^n)$ is $F_0 = (e_1, ..., e_m)$ (the first $m$ members of the standard orthonormal basis of $\mathbb{K}^n$ listed in natural order). Its stabilizer, $H$, equals the subgroup of all matrices in $GL_n(\mathbb{K})$ whose first $m$ rows equal $e_1, e_2, ..., e_m$ (in that order). Lemma* 10.17 *provides a natural identification $S_m(\mathbb{K}^n) \cong H\backslash GL_n(\mathbb{K})$. Since this coset space has the structure of a homogeneous manifold, the Steifel manifold inherits the same via this identification.*

The Grassmann and Steifel manifolds are previously unfamiliar sets that can be given the structure of homogeneous manifolds by identifying transitive actions on them. We now turn our attention to a smooth transitive action on a set $M$ that already has a manifold structure. In this case, we wish to verify that the manifold structure that $M$ inherits via the identification with a coset space is the same as its given manifold structure. More significantly, we wish to prove that every homogeneous manifold is a coset space. These wishes are granted by the following smooth version of Lemma 10.17:

**Theorem 10.19.** *In Lemma* 10.17*, if $M$ is a manifold and $G$ is a path connected matrix group and $\varphi$ is smooth, then $F$ is a diffeomorphism.*

We'll see in the proof that the "path-connected" hypothesis can be replaced with the weaker hypothesis that $G$ has only countably many connected components.

**Sketch of proof.** We know from Lemma 10.17 that $F$ is a well-defined bijection, while Proposition 10.10 implies that $F$ is smooth. According to the Inverse Function Theorem for generalized manifolds (Exercise 10.4), it will suffice to prove that for any $g \in G$, the linear transformation $dF_{[g]} : T_{[g]}(G/H) \to T_{g\star p_0}M$ is an isomorphism.

In fact, it is enough to verify the special case when $g = I$. Why? Because the diagram in Equation 10.5 gives that for any $g \in G$,

$$F = \underbrace{\varphi(g)}_{\text{diffeo}} \circ F \circ \underbrace{\tilde{\varphi}(g)^{-1}}_{\text{diffeo}}.$$

So the chain rule for generalized manifolds (Exercise 10.3) gives:

$$dF_{[g]} = \underbrace{d\left(\varphi(g)\right)_{p_0}}_{\text{isomorphism}} \circ dF_{[I]} \circ \underbrace{d\left(\tilde{\varphi}(g)^{-1}\right)_{[g]}}_{\text{isomorphism}}.$$

Thus, $dF_{[g]}$ is an isomorphism if and only if $dF_{[I]}$ is an isomorphism.

We will next prove that $dF_{[I]} : T_{[I]}(G/H) \to T_{p_0}M$ is injective (has trivial kernel). Suppose to the contrary that $dF_{[I]}(X) = 0$ for some non-zero $X \in T_{[I]}(G/H)$. Let $\mathfrak{h}$ denote the Lie algebra of $H$ and define $\mathfrak{p} = \mathfrak{h}^{\perp}$. Notice that

$$0 = dF_{[I]}(X) = \frac{d}{dt}\Big|_{t=0} \exp(tX) \star p_0,$$

where on the right, we are regarding $X$ as an element of $\mathfrak{p}$ via the natural identification $T_{[I]}(G/H) \cong \mathfrak{p}$ provided by Theorem 10.8. For all $t_0 \in \mathbb{R}$, we have:

$$\begin{aligned} \frac{d}{dt}\Big|_{t=t_0} \exp(tX) \star p_0 &= \frac{d}{dt}\Big|_{t=0} \exp((t_0+t)X) \star p_0 \\ &= \frac{d}{dt}\Big|_{t=0} (\exp(t_0X)\cdot\exp(tX)) \star p_0 \\ &= \frac{d}{dt}\Big|_{t=0} \varphi\left(\exp(t_0X)\right)\left(\exp(tX)\star p_0\right) \\ &= d\left(\varphi(\exp(t_0X))\right)_{p_0}\left(\frac{d}{dt}\Big|_{t=0} \exp(tX)\star p_0\right) = 0. \end{aligned}$$

Since the path $t \mapsto \exp(tX) \star p_0$ has zero derivative for all time, it is constant: $\exp(tX) \star p_0 = p_0$ for all $t \in \mathbb{R}$. Therefore $\exp(tX) \in H$ for all $t \in \mathbb{R}$. This implies that $X \in \mathfrak{h}$, which contradicts the hypothesis that $X$ is a non-zero vector in $\mathfrak{p}$.

In summary, $dF_{[I]} : T_{[I]}(G/H) \to T_{p_0}M$ is injective, which would force it to also be surjective if we knew that $\dim(G/H) = \dim(M)$. It therefore remains to rule out the possibility that $M$ has a larger dimension that $G/H$. This requires some topology, and we refer the reader to [**11**] or [**13**] for a proof. The path-connected hypothesis is used here, but can be replaced by the weaker hypothesis that $G$ has countably many connected components. In fact, this hypothesis can be completely removed, provided one follows the common convention of adding two topological conditions to the definition of manifold (namely, "Hausdorff" and "second countable"). $\square$

The previous theorem allows us to better understand many familiar manifolds by identifying them with coset spaces. To give an example, we must first adopt some notation conventions. First, we will always think of products of matrix groups as having the most obvious block-diagonal form, so if $H_1 \subset GL_{n_1}(\mathbb{K})$ and $H_2 \subset GL_{n_2}(\mathbb{K})$, then "$H_1 \times H_2$" will denote the following subgroup of $GL_{n_1+n_2}(\mathbb{K})$:

$$H_1 \times H_2 = \{\mathrm{diag}(A, B) \mid A \in H_1, B \in H_2\}.$$

Further, when $H \subset GL_n(\mathbb{K})$ is a subgroup, we will denote:

$$S(H) = \{M \in H \mid \det(M) = 1\}.$$

For example, $S(O(n)) = SO(n)$, $S(U(n)) = SU(n)$ and $S(Sp(n)) = Sp(n)$. With this notation, we have:

**Corollary 10.20.** *In each of the following, $n \geq 1$ and "$=$" means the manifold is diffeomorphic to the coset space:*

(1) $$\mathbb{KP}^n = \mathcal{O}_{n+1}(\mathbb{K})/\left(\mathcal{O}_1(\mathbb{K}) \times \mathcal{O}_n(\mathbb{K})\right) = S\left(\mathcal{O}_{n+1}(\mathbb{K})\right)/S\left(\mathcal{O}_1(\mathbb{K}) \times \mathcal{O}_n(\mathbb{K})\right)$$

*In particular:*

$$\begin{aligned} \mathbb{RP}^n &= O(n+1)/\left(\{\pm 1\} \times O(n)\right) \\ &= SO(n+1)/S\left(\{\pm 1\} \times O(n)\right) \\ \mathbb{CP}^n &= U(n+1)/\left(U(1) \times U(n)\right) \\ &= SU(n+1)/S\left(U(1) \times U(n)\right) \\ \mathbb{HP}^n &= Sp(n+1)/\left(Sp(1) \times Sp(n)\right) \end{aligned}$$

(2) $$S^n = O(n+1)/(\{1\} \times O(n)) = SO(n+1)/\left(\{1\} \times SO(n)\right)$$

(3) $$S^{2n+1} = U(n+1)/\left(\{1\} \times U(n)\right) = SU(n+1)/\left(\{1\} \times SU(n)\right)$$

(4) $$S^{4n+3} = Sp(n+1)/\left(\{1\} \times Sp(n)\right)$$

**Proof.** In each example, written as "$M = G/H$," we must identify a smooth transitive action of $G$ on $M$ and a point of $M$ whose stabilizer equals $H$. Since $G/H$ is diffeomorphic to $H\backslash G$, it doesn't matter whether we identify a left or right action. In all cases, the action is the standard action of $\mathcal{O}_{n+1}(\mathbb{K})$ (or a subgroup thereof) on $\mathbb{K}^{n+1}$ (or the induced action on the lines in $\mathbb{K}^{n+1}$ or the induced action on the sphere of unit-length vectors in $\mathbb{K}^{n+1}$) for the appropriate choice of $\mathbb{K}$. Furthermore, in all cases, $p_0$ is either $e_1$ or $[e_1]$, where $e_1 = (1, 0, ..., 0)$ denotes the first member of the standard orthonormal basis of the $\mathbb{K}^{n+1}$. The transitivity of these actions follows from Exercise 3.18. □

## 6. Riemannian manifolds

Since a manifold is locally identified with Euclidean space, one can do to a manifold *almost* everything that one can do to Euclidean space. But there is a crucial exception: one cannot measure lengths of curves or distances between pairs of points. Attempting to do so using parametrizations would lead to answers that would depend on the choice of parametrization and therefore would not be well-defined.

These “metric” measurements require an additional structure that is specified in the following two definitions.

**Definition 10.21.** *Let $\mathcal{V}$ be a finite-dimensional vector space over $\mathbb{R}$. An **inner product** on $\mathcal{V}$ is a bilinear function $\mathcal{V} \times \mathcal{V} \to \mathbb{R}$, denoted as $v, w \mapsto \langle v, w \rangle$, satisfying the following properties:*

(1) (symmetric) $\langle v, w \rangle = \langle w, v \rangle$ *for all* $v, w \in \mathcal{V}$.

(2) (positive definite) $\langle v, v \rangle \geq 0$ *for all* $v \in \mathcal{V}$, *with equality if and only if* $v = 0$.

In other words, an inner product is a function that has the same algebraic properties of the standard inner product on $\mathbb{R}^n$, as enumerated in Section 1 of Chapter 3. Just as with the standard inner product, an arbitrary inner product empowers one to compute norms and angles as follows:

$$|v| = \sqrt{\langle v, v \rangle} \quad \text{and} \quad \angle(v, w) = \arccos\left(\frac{\langle v, w \rangle}{|v||w|}\right).$$

These definitions generalize Definition 3.2 and Equation 3.4 respectively.

If $\mathcal{V}$ is a subspace of $\mathbb{R}^n$ for some $n$, then the standard inner product on $\mathbb{R}^n$ restricts to an inner product on $\mathcal{V}$. But this is not the only inner product on $\mathcal{V}$. In fact, given any basis of $\mathcal{V}$, there exists a unique inner product on $\mathcal{V}$ with respect to which this basis is orthonormal (Exercise 10.26). Different inner products yield different calculations of norms and angles.

**Definition 10.22.** *Let $M$ be a manifold. A **Riemannian metric** on $M$ is a choice for each $p \in M$ of an inner product, $\langle \cdot, \cdot \rangle_p$, on $T_pM$ that “varies smoothly with $p$” in the following sense: for any pair, $X$ and $Y$, of smooth vector fields on $M$, the map $p \mapsto \langle X(p), Y(p) \rangle_p$ is a smooth function from $M$ to $\mathbb{R}$. A manifold together with a Riemannian metric is called a **Riemannian manifold**.*

A Riemannian metric on $M$ empowers one to compute norms of (and angles between) its tangent vectors at any point. For example, the norm of $v \in T_pM$ is defined as $|v|_p = \sqrt{\langle v, v \rangle_p}$. Building on this, one can define the **length** of a smooth path $\gamma : [a, b] \to M$ to equal

$\int_a^b |\gamma'(t)|_{\gamma(t)}\,\mathrm{dt}$. The **distance** between a pair of points $p, q \in M$ can then be defined as the infimum length of all smooth paths $\gamma : [a, b] \to M$ with $\gamma(a) = p$ and $\gamma(b) = q$. When $M$ is path-connected, this infimum is always finite, and this distance function makes $M$ into a metric space. The topology induced by this distance function is the same as the topology described in Definition 10.2. The reader will prove these assertions in Exercise 10.27.

The natural notion of equivalence for Riemannian manifolds is:

**Definition 10.23.** *A diffeomorphism $f : M_1 \to M_2$ between Riemannian manifolds is called an **isometry** if its derivative preserves inner products; that is, for all $p \in M_1$ and all $v, w \in T_pM_1$, we have*

$$\langle df_p(v), df_p(w)\rangle_{f(p)} = \langle v, w\rangle_p.$$

*Two Riemannian manifolds are called **isometric** if there exists an isometry between them.*

A diffeomorphism between Riemannian manifolds is an isometry if and only if it preserves distances (Exercise 10.28), so this definition is consistent with the common use of the word "isometry" within the study of metric spaces.

If $M \subset \mathbb{R}^m$ is an embedded manifold, then each of its tangent spaces is a subspace of $\mathbb{R}^m$. The natural **embedded Riemannian metric** on $M$ just means the restriction to each tangent space of the standard inner product on $\mathbb{R}^m$. As previously mentioned, every matrix group $G \subset GL_n(\mathbb{K}) \subset M_n(\mathbb{K})$ is an embedded manifold and therefore inherits a natural embedded Riemannian metric. When $G \subset \mathcal{O}_n(\mathbb{K})$, this embedded Riemannian metric has special symmetry properties that were studied in Section 4 of Chapter 8.

The **Nash Embedding Theorem** states that every Riemannian manifold is isometric to an embedded manifold with the embedded Riemannian metric. This means that the class of generalized manifolds with arbitrary Riemannian metrics is really no more general than the class of embedded manifolds with embedded metrics. However, some of the Riemannian manifolds we will describe in this section would not look elegant or simple if one attempted to describe them as embedded.

Our next goal is to describe a natural Riemannian metric on certain coset spaces of the form $G/H$. The metric on $G/H$ will be contrived such that the projection map $\pi : G \to G/H$ preserves the Riemannian metrics. What should this mean? A smooth function from a higher-dimensional Riemannian manifold to a lower-dimensional Riemannian manifold could never be an isometry, yet there is a meaningful sense in which it might preserve the Riemannian metrics:

**Definition 10.24.** *Let $f : M \to B$ be a smooth function between Riemannian manifolds with $k = dim(M) - dim(B) > 0$. Then $f$ is called a **Riemannian submersion** if for all $p \in M$:*

(1) *$df_p : T_pM \to T_{f(p)}B$ is surjective (which implies that the kernel of $df_p$ has dimension $k$).*
(2) *For all $v, w \in T_pM$ orthogonal to the kernel of $df_p$,*

$$\langle df_p(v), df_p(w)\rangle_{f(p)} = \langle v, w\rangle_p.$$

**Theorem 10.25.** *If $G \subset \mathcal{O}_n(\mathbb{K})$ is a matrix group and $H \subset G$ is a closed subgroup, there exists a unique Riemannian metric on $G/H$ (called the **submersion metric**) such that the projection map $\pi : G \to G/H$ is a Riemannian submersion.*

The distance function induced by this Riemannian metric on $G/H$ is quite natural. It can be shown that the distance between a pair of cosets in $G/H$ can be computed by regarding the cosets as subsets of $G$ and calculating the infimum distance within $G$ between a point of the first coset and a point of the second.

**Proof.** Let $g \in G$ be an arbitrary element, so $[g] \in G/H$ is an arbitrary coset. The parametrization $\varphi_g$ defined in Theorem 10.8 provides a natural identification $F_g : \mathfrak{p} \to T_{[g]}(G/H)$, namely:

$$F_g(X) = \frac{d}{dt}\Big|_{t=0} [g \cdot \exp(tX)] \tag{10.6}$$

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

As in Section 4 of Chapter 8, we will denote by $\langle\cdot,\cdot\rangle_{\mathbb{R}}$ the inner product on $\mathfrak{g}$ coming from the embedded Riemannian metric on $G$. We will use the same notation for the restriction of this inner product to the subspace $\mathfrak{p} \subset \mathfrak{g}$. Via the identification $F_g$, this inner product on

$\mathfrak{p}$ naturally induces an inner product $T_{[g]}(G/H)$. In other words, an arbitrary pair of vectors in $T_{[g]}(G/H)$ can be written as $F_g(X), F_g(Y)$ for some $X, Y \in \mathfrak{p}$, and we define:

$$\langle F_g(X), F_g(Y)\rangle_{[g]} = \langle X, Y\rangle_{\mathbb{R}}.$$

We must prove that the inner product on $T_{[g]}(G/H)$ constructed in this manner does not depend on the choice of representative, $g$, of the coset $[g]$. An alternative representative would have the form $g \cdot h$ for some $h \in H$. It is straightforward to verify that for all $X \in \mathfrak{p}$:

$$F_{gh}(X) = F_g(\mathrm{Ad}_h X).$$

So the inner product is well-defined because for all $X, Y \in \mathfrak{p}$,

$$\begin{aligned}\langle F_{gh}(X), F_{gh}(Y)\rangle_{[gh]} &= \langle X, Y\rangle_{\mathbb{R}} \\ &= \langle \mathrm{Ad}_h X, \mathrm{Ad}_h Y\rangle_{\mathbb{R}} = \langle F_g(\mathrm{Ad}_h X), F_g(\mathrm{Ad}_h Y)\rangle_{[g]}.\end{aligned}$$

The second equality above reflects the Ad-invariant property of $\langle\cdot,\cdot\rangle_{\mathbb{R}}$ that was described in Section 4 of Chapter 8.

We leave to the reader in Exercise 10.29 the verification that this Riemannian metric varies smoothly with the point and that $\pi : G \to G/H$ is a Riemannian submersion. $\square$

We end this section by describing the natural symmetry property of the submersion metrics constructed in the previous proof. For this, we must first specify what it means for an action to respect a Riemannian metric:

**Definition 10.26.** *A smooth action, $\varphi$, of a matrix group $G$ on a Riemannian manifold $M$ is called an* ***isometric action*** *if for every $g \in G$, the diffeomorphism $\varphi(g) : M \to M$ is an isometry. A Riemannian manifold, $M$, is called a* ***Riemannian homogeneous manifold*** *if there exists a transitive smooth isometric action of a matrix group on $M$.*

The Riemannian version of Example 10.14 is:

**Proposition 10.27.** *Under the assumptions of Theorem* 10.25*, the left action of $G$ on $G/H$ is an isometric action, so $G/H$ is a Riemannian homogeneous manifold (with respect to the submersion metric).*

**Proof.** Let $g \in G$ and denote by $\mathcal{L}_g : (G/H) \to (G/H)$ the left-multiplication map, defined as $\mathcal{L}_g([x]) = [g \cdot x]$. We must prove that for all $x \in G$, the linear transformation

$$d(\mathcal{L}_g)_{[x]} : T_{[x]}(G/H) \to T_{[gx]}(G/H)$$

preserves inner products. As in the previous proof, $\mathfrak{p}$ is identified with both of these tangent spaces such that the following diagram commutes:

$$\begin{array}{ccc} \mathfrak{p} & \xrightarrow{\text{identity}} & \mathfrak{p} \\ {\scriptstyle F_x}\downarrow & & \downarrow{\scriptstyle F_{gx}} \\ T_{[x]}(G/H) & \xrightarrow{d(\mathcal{L}_g)_{[x]}} & T_{[gx]}(G/H) \end{array} \tag{10.7}$$

Since the inner products on these tangent spaces are defined via these identifications, it follows that $d(\mathcal{L}_g)_{[x]}$ preserves inner products. □

Combining these results with Corollary 10.20 provides the structure of a Riemannian homogeneous manifold on all of the projective spaces: $\mathbb{RP}^n$, $\mathbb{CP}^n$, and $\mathbb{HP}^n$. Spheres and projective spaces are among the most natural Riemannian manifolds.

## 7. Lie groups

Essentially all of the theory of matrix groups in this book is also true for a generalization of matrix groups called *Lie groups*. In this section, we briefly overview (without proofs) the structures and theorems for matrix groups that carry over to Lie groups. We begin with the definition:

**Definition 10.28.** *A* ***Lie group*** *is a manifold, $G$, with a smooth group operation $G \times G \to G$.*

In other words, a Lie group is a manifold that is also a group. Many authors add to this definition the requirement that the "inverse map" $G \to G$, sending $g \mapsto g^{-1}$, is smooth; however, this turns out to be a consequence of the smoothness of the group operation $(g_1, g_2) \mapsto g_1 \cdot g_2$.

In Chapter 7, we proved that matrix groups are (embedded) manifolds. The group operation is smooth, so *matrix groups are Lie groups.*

All important structures of matrix groups carry over to Lie groups. For example, the Lie algebra, $\mathfrak{g}$, of a Lie group $G$ is defined as you would expect:

$$\mathfrak{g} = T_I G.$$

For every $g \in G$, the conjugation map $C_g : G \to G$ sending $x \mapsto gxg^{-1}$ is smooth, so one can define:

$$Ad_g = d(C_g)_I : \mathfrak{g} \to \mathfrak{g}.$$

Next, the Lie bracket operation in $\mathfrak{g}$ is defined as you would expect: for $A, B \in \mathfrak{g}$,

$$[A, B] = \frac{d}{dt}\Big|_{t=0} \mathrm{Ad}_{a(t)} B,$$

where $a(t)$ is any differentiable path in $G$ with $a(0) = I$ and with $a'(0) = A$. It turns out that this operation satisfies the familiar Lie bracket properties of Proposition 8.4. Next, the exponential map

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

is defined with inspiration from Proposition 6.10: For $A \in \mathfrak{g}$, the path $t \mapsto e^{tA}$ (at least for sufficiently small $t$) means the integral curve of the vector field on $G$ whose value at $g \in G$ is $d(\mathcal{L}_g)_I(A) \in T_g G$, where $\mathcal{L}_g : G \to G$ denotes the map $x \mapsto g \cdot x$.

Moreover, every *compact* Lie group has a maximal torus, and the key facts from Chapter 9 generalize to this setting. In particular:

**Proposition 10.29.** *Let $G$ be a path-connected compact Lie group, let $T$ be a maximal torus of $G$, and let $\tau \subset \mathfrak{g}$ denote the Lie algebras of $T \subset G$.*

(1) *For every $x \in G$ there exists $g \in G$ such that $x \in g \cdot T \cdot g^{-1}$.*

(2) *Every maximal torus of $G$ equals $g \cdot T \cdot g^{-1}$ for some $g \in G$.*

(3) *If $x \in G$ commutes with every element of $T$, then $x \in T$.*

(4) *If $X \in \tau$ commutes with every element of $\tau$, then $X \in \tau$.*

In Chapter 9, we proved these claims for $SO(n)$, $U(n)$, $SU(n)$ and $Sp(n)$, with different proofs for each case. The proof techniques required for general compact Lie groups (or even general compact matrix groups) are a bit beyond the scope of this book.

Our previous results about coset spaces generalize to Lie groups. In particular, if $G$ is a Lie group and $H \subset G$ is a closed subgroup, then the coset space $G/H$ has a natural manifold structure with respect to which the left action of $G$ on $G/H$, defined as $g\star[x] = [g\cdot x]$, is smooth and transitive. Conversely, if a path-connected Lie group $G$ acts smoothly and transitively on a manifold $M$, then $M$ is diffeomorphic to $G/G_{p_0}$ for any $p_0 \in M$.

Lie groups are among the most fundamental objects in mathematics and physics, and the most significant Lie groups are matrix groups. In fact, Lie groups turn out to be only *slightly* more general than matrix groups. One indication of this is Ado's Theorem: the Lie algebra of any Lie group is isomorphic to the Lie algebra of a matrix group. Another indication is that every *compact* Lie group turns out to be smoothly isomorphic to a matrix group.

In fact, a major achievement of modern mathematics is the classification of compact Lie groups. The only such groups we have encountered so far are $SO(n)$, $O(n)$, $U(n)$, $SU(n)$, $Sp(n)$, and products of these, such as, for example, $SO(3) \times SO(5) \times SU(2)$. It turns out that there are not many more than these.

**Theorem 10.30.** *The Lie algebra of any compact Lie group is isomorphic to the Lie algebra of a product $G_1 \times G_2 \times \cdots \times G_k$, where each $G_i$ is one of $\{SO(n), SU(n), Sp(n)\}$ for some $n$ or is one of a list of five possible exceptions.*

The five "exceptional Lie groups" mentioned in the theorem are named:

$G_2$, which has dimension 14

$F_4$, which has dimension 54

$E_6$, which has dimension 78

$E_7$, which has dimension 133

$E_8$, which has dimension 248

We have seen that non-isomorphic Lie groups sometimes have isomorphic Lie algebras. For example, $U(n)$ is not on the list in Theorem 10.30 because it has the same Lie algebra as $SU(n)\times SO(2)$, by Exercise 4.21. The problem of determining all Lie groups with the same Lie algebra as $G_1 \times G_2 \times \cdots \times G_k$ is well-understood, but is also beyond the scope of this text. Aside from this detail, the theorem gives a complete classification of compact Lie groups!

Although we will not construct the exceptional Lie groups (which are all matrix groups) or fully prove Theorem 10.30, the next chapter will overview the key ideas of the proof. To pave the way, we will end this chapter with a key result: every compact Lie group $G$ has a **bi-invariant metric**, which means a Riemannian metric for which the left and right actions of $G$ on $G$ are both isometric actions.

Although we will not fully prove this fact, we will describe how to reduce it to a more tractable assertion. Describing a Riemannian metric on a general manifold requires a lot of information: a different inner product must be chosen for each tangent space. Directly managing this much information is generally intractable. But on a Lie group, one can describe a metric simply by choosing an inner product on a *single* vector space, namely its Lie algebra:

**Definition 10.31.** *Let $G$ be a Lie group with Lie algebra denoted $\mathfrak{g}$. Any inner product, $\langle \cdot,\cdot\rangle_I$, on $\mathfrak{g}$ determines a Riemannian metric on $G$ in either of the following ways:*

(1) ***(Left-invariant metric)*** *For all $g \in G$ and $v, w \in T_gG$, define:*

$$\langle v, w\rangle_g = \langle d(\mathcal{L}_{g^{-1}})_g(v), d(\mathcal{L}_{g^{-1}})_g(v)\rangle_I,$$

*where $\mathcal{L}_g : G \to G$ is defined as $\mathcal{L}_g(x) = g\cdot x$.*

(2) ***(Right-invariant metric)*** *For all $g \in G$ and $v, w \in T_gG$, define:*

$$\langle v, w\rangle_g = \langle d(\mathcal{R}_{g^{-1}})_g(v), d(\mathcal{R}_{g^{-1}})_g(v)\rangle_I,$$

*where $\mathcal{R}_g : G \to G$ is defined as $\mathcal{R}_g(x) = x\cdot g$.*

The idea is that for each $g \in G$, $\mathfrak{g}$ is naturally identified with $T_gG$ via either $d(\mathcal{L}_g)_I$ or $d(\mathcal{R}_g)_I$. Via either of these identifications, an inner product on $\mathfrak{g}$ induces an inner product on all other tangent spaces. These two choices are named after the symmetries they possess. With respect to a left-invariant metric on a Lie group $G$, the left action of $G$ on $G$ is isometric. Similarly, with respect to a right-invariant metric, the right action of $G$ on $G$ is isometric.

For a *compact* Lie group, we don't have to choose:

**Theorem 10.32.** *If $G$ is a compact Lie group with Lie algebra denoted $\mathfrak{g}$, then there exists an inner product, $\langle\cdot,\cdot\rangle$, on $\mathfrak{g}$ satisfying the following equivalent conditions:*

(1) *(**Ad-invariance**) For all $g \in G$ and all $v, w \in \mathfrak{g}$,*

$$\langle Ad_g(v), Ad_g(w)\rangle = \langle v, w\rangle.$$

(2) *(**Bi-invariance**) The left-invariant Riemannian metric on $G$ determined by $\langle\cdot,\cdot\rangle$ is the same as the right-invariant Riemannian metric (and is therefore called a **bi-invariant metric**).*

The equivalence of (1) and (2) is straightforward (Exercise 10.31), but it requires some new ideas to prove that an Ad-invariant inner product exists. We will not address the proof here.

For the compact matrix group $\mathcal{O}_n(\mathbb{K})$, the inner product inherited from the ambient Euclidean space, previously denoted $\langle\cdot,\cdot\rangle_{\mathbb{R}}$, was proven in Section 4 of Chapter 8 to be Ad-invariant by first re-describing it as: $\langle X, Y\rangle_{\mathbb{R}} = \text{Real}(X \cdot Y^*)$. Thus, Theorem 10.32 generalizes a familiar structure on $\mathcal{O}_n(\mathbb{K})$ to arbitrary compact Lie groups.

We end this section by stating without proof a powerful generalization of Theorem 10.8 and Theorem 10.25:

**Theorem 10.33.** *Given a smooth free isometric left action of a compact Lie group $G$ on a Riemannian manifold $M$, the orbit space, $G\backslash M$, has a unique manifold structure and Riemannian metric such that the projection $\pi : M \to G\backslash M$ is a smooth Riemannian submersion.*

## 8. Exercises

**Ex. 10.1.** Prove that an embedded manifold (Definition 7.14) is a manifold (Definition 10.1).

**Ex. 10.2.** Prove the assertion in Section 1 that an atlas satisfying conditions (1) and (2) of Definition 10.1 can be uniquely completed to an atlas that also satisfies condition (3).

**Ex. 10.3.** Prove the chain rule (Proposition 7.21) for general (not necessarily embedded) manifolds.

**Ex. 10.4.** Prove the Inverse Function Theorem (Theorem 7.22) for general (not necessarily embedded) manifolds.

**Ex. 10.5.** If $M_1$ and $M_2$ are (generalized) manifolds, describe a natural manifold structure on their product, $M_1 \times M_2$, such that

$$\dim(M_1 \times M_2) = \dim(M_1) + \dim(M_2).$$

Prove that the projection $f_1 : M_1 \times M_2 \to M_1$, defined as $f(p_1, p_2) = p_1$, is smooth, as is the analogously defined projection $f_2$.

**Ex. 10.6.** State and prove generalizations of Exercises 7.11, 7.12 and 7.13 for general (not necessarily embedded) manifolds.

**Ex. 10.7.** Let $X$ be a smooth vector field on a manifold, $M$. Prove that for *any* parametrization, $\varphi : U \subset \mathbb{R}^d \to V \subset M$, the naturally associated function from $U$ to $\mathbb{R}^d$ is smooth.

**Ex. 10.8.** Prove that $\mathbb{RP}^1$ is diffeomorphic to $SO(2)$.

**Ex. 10.9.** Prove the assertion in Proposition 10.6 that $\mathbb{HP}^1$ is diffeomorphic to $S^4$.

**Ex. 10.10.** Prove Proposition 10.10.

**Ex. 10.11.** Let $H \subset K \subset G$ be nested closed matrix groups. Prove that the natural map $f : G/H \to G/K$, defined as $f(g \cdot H) = g \cdot K$, is smooth and that the preimage of any element of $G/K$ is diffeomorphic to $K/H$.

**Ex. 10.12.** Let $\varphi$ be a right action of a group $G$ on a set $M$. Define $\overline{\varphi}(g) = \varphi(g^{-1})$ for all $g \in G$. Prove that $\overline{\varphi}$ is a left action of $G$ on $M$.

**Ex. 10.13.** For a transitive action of a group $G$ on a set $M$, prove that any two stabilizers are conjugate.

**Ex. 10.14.** For a smooth action of a matrix group $G$ on a manifold $M$, prove that all stabilizers are *closed* subgroups of $G$.

**Ex. 10.15.** Prove Lemma 10.17.

**Ex. 10.16.** In Example 10.14, prove the assertion that the natural left action of $G$ on $G/H$ is smooth.

**Ex. 10.17.** For positive integers $m < n$, prove that the Grassmann manifolds $G_m(\mathbb{K}^n)$ and $G_{n-m}(\mathbb{K}^n)$ are diffeomorphic.

**Ex. 10.18.** With notation as in Corollary 10.20, find $H$ such that

$$\mathbb{KP}^n = GL_n(\mathbb{K})/H = SL_n(\mathbb{K})/S(H).$$

**Ex. 10.19.** With notation as in Corollary 10.20, find $H$ such that

$$T^1S^n = O(n+1)/H$$

(see Exercise 7.13 for the definition of the unit tangent bundle).

**Ex. 10.20.** With notation as in Corollary 10.20, show that:

$$G_m(\mathbb{R}^n) = O(n)/\left(O(m) \times O(n-m)\right).$$

**Ex. 10.21** (Flag manifolds)**.** A **flag** in $\mathbb{C}^n$ means a collection of nested $\mathbb{C}$-subspaces:

$$V_1 \subset V_2 \subset \cdots \subset V_{n-1}$$

such that $\dim(V_i) = i$ for each $i$. Let $F$ denote the set of all flags in $\mathbb{C}^n$. Verify that the left action of $SU(n)$ on $F$ defined as

$$g \star (V_1 \subset V_2 \subset \cdots \subset V_{n-1}) = (g \cdot V_1 \subset g \cdot V_2 \subset \cdots \subset g \cdot V_{n-1})$$

is transitive and that every stabilizer is a maximal torus of $SU(n)$.

**Ex. 10.22.** One of the $n = 1$ cases of Corollary 10.20 says that $\mathbb{CP}^1 = SU(2)/S(U(1)\times U(1))$. How is this transitive action of $SU(2)$ on $\mathbb{CP}^1$ related to the adjoint action of $Sp(1)$ on $S^2$ whereby the double cover $Sp(1) \to SO(3)$ was constructed in Section 6 of Chapter 8?

**Ex. 10.23.** Prove that $S(U(1) \times U(n))$ is smoothly isomorphic to $U(n)$ and that $S(\{\pm 1\} \times O(n))$ is smoothly isomorphic to $O(n)$. *Comment: These groups are mentioned in Corollary* 10.20.

**Ex. 10.24.** For a homogeneous manifold $M = G/H$, the parametrization $\varphi_I$ defined in Theorem 10.8 provides a natural identification of $\mathfrak{p}$ with $T_{[I]}(G/H)$ (this identification was named $F_I$ in Equation 10.6). Explicitly describe $\mathfrak{p}$ for the following examples:

(1) $S^4 = SO(5)/(\{1\} \times SO(4))$.
(2) $\mathbb{CP}^2 = U(3)/(U(1) \times U(2))$.
(3) $\mathbb{CP}^2 = SU(3)/S(U(1) \times U(2))$.
(4) $G_2(\mathbb{R}^5) = GL_5(\mathbb{R})/H_0$ as in Equation 10.4.
(5) $G_2(\mathbb{R}^5) = O(5)/\,(O(2) \times O(3))$ as in Exercise 10.20.
(6) $S_2(\mathbb{C}^5) = GL_5(\mathbb{C})/H$ as in Example 10.18.

**Ex. 10.25.** Let $\{e_1, e_2, e_3, e_4, e_5\}$ denote the standard orthonormal basis of $\mathbb{R}^5$. Let $V_0 = \text{span}\{e_1, e_2\} \subset \mathbb{R}^5$, so $V_0^\perp = \text{span}\{e_3, e_4, e_5\}$. Let $t \mapsto f(t)$ be a differentiable one-parameter family of linear transformations from $V_0$ to $V_0^\perp$, which can be expressed with respect to the above bases as a one-parameter family of matrices:

$$f(t) = \begin{pmatrix} f_{11}(t) & f_{12}(t) & f_{13}(t) \\ f_{21}(t) & f_{22}(t) & f_{23}(t) \end{pmatrix},$$

where each $f_{ij}$ is a differentiable function with $f_{ij}(0) = 0$. Define:

$$V(t) = \text{the graph of } f(t) = \{v + f(t)(v) \mid v \in V_0\},$$

so that $t \mapsto V(t)$ is a path in $G_2(\mathbb{R}^5)$ whose initial derivative, $V'(0) \in T_{V_0}(G_2(\mathbb{R}^5))$, is determined by the six elements of $f'(0)$. How is this related to the six real numbers used to describe $\mathfrak{p}$ in parts (4) or (5) of Exercise 10.24?

**Ex. 10.26.** Let $\mathcal{V}$ be a finite-dimensional vector space over $\mathbb{R}$. Given any basis of $\mathcal{V}$, prove there exists a unique inner product on $\mathcal{V}$ with respect to which the basis is orthonormal.

**Ex. 10.27.** If $M$ is a path-connected Riemannian manifold, prove that the distance between any pair of points is finite and that the distance function makes $M$ into a metric space. Prove that the topology induced by this distance function is the same as the topology described in Definition 10.2.

**Ex. 10.28.** Prove that a diffeomorphism between Riemannian manifolds is an isometry if and only if it preserves distances.

**Ex. 10.29.** Complete the proof of Theorem 10.25 by proving that the Riemannian metric varies smoothly with the point and that $\pi : G \to G/H$ is a Riemannian submersion.

**Ex. 10.30.** If $G$ is a matrix group and $H \subset G$ is a closed *normal* subgroup, prove that $G/H$ is a Lie group.

**Ex. 10.31.** In Theorem 10.32, prove that (1) and (2) are equivalent.

**Ex. 10.32.** Prove that an Ad-invariant inner product, $\langle \cdot, \cdot \rangle$, on the Lie algebra of a Lie group $G$ must satisfy the following **infinitesimal Ad-invariance** property: for all $A, B, C \in \mathfrak{g}$,

$$\langle [A, B], C \rangle = -\langle [A, C], B \rangle.$$

*Hint: Copy the proof of Proposition* 8.14*.*

**Ex. 10.33** (An exotic sphere)**.** Consider the smooth left action of the matrix group $H = Sp(1) \times Sp(1)$ on the manifold $G = Sp(2)$ defined so that for all $q_1, q_2 \in Sp(1)$ and all $A \in Sp(2)$:

$$(q_1, q_2) \star A = \begin{pmatrix} q_1 & 0 \\ 0 & q_1 \end{pmatrix} \cdot A \cdot \begin{pmatrix} \overline{q_2} & 0 \\ 0 & 1 \end{pmatrix}.$$

Prove that this action is free.

*Comment: Theorem* 10.33 *provides the orbit space with the structure of a* 7*-dimensional Riemannian manifold. This orbit space has been proven to be an* ***exotic sphere****: it is homeomorphic but not diffeomorphic to* $S^7$*.*

# Chapter 11

# Roots

According to Theorem 10.30, the "classical" compact Lie groups, $SO(n)$, $SU(n)$, and $Sp(n)$, together with the five exceptional groups, form the building blocks of all compact Lie groups. In this chapter, we will use roots to better understand the Lie bracket operation in the Lie algebra, $\mathfrak{g}$, of a classical or general compact Lie group $G$.

For each of the classical compact Lie groups, we will describe a natural basis for the Lie algebra, contrived so that the Lie bracket of any pair of basis elements is particularly simple (in most cases, it is a multiple of another basis element). Roots are used to describe in a uniform way the patterns governing how basis elements bracket together. They address the question: what do the bracket structures of $so(n)$, $su(n)$ and $sp(n)$ (and perhaps even the Lie algebra of an *arbitrary* compact Lie group) have in common?

This chapter is organized as follows. First we will explicitly describe the roots and the bracket operation for $SU(n)$. Next we will define and study the roots of an arbitrary compact Lie group. We will then apply this general theory to describe the roots and the bracket operation for $SO(n)$ and $Sp(n)$. Finally, we will roughly indicate how the theory of roots leads to a proof of Theorem 10.30.

## 1. The structure of $su(3)$

Consider the following basis for $su(3)$:

$$H_{12} = \begin{pmatrix} \mathbf{i} & 0 & 0 \\ 0 & -\mathbf{i} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad E_{12} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad E_{23} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix},$$

$$E_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad H_{23} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mathbf{i} & 0 \\ 0 & 0 & -\mathbf{i} \end{pmatrix}, \quad F_{12} = \begin{pmatrix} 0 & \mathbf{i} & 0 \\ \mathbf{i} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$F_{23} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \mathbf{i} \\ 0 & \mathbf{i} & 0 \end{pmatrix}, \quad F_{13} = \begin{pmatrix} 0 & 0 & \mathbf{i} \\ 0 & 0 & 0 \\ \mathbf{i} & 0 & 0 \end{pmatrix}.$$

The Lie bracket of any pair of these basis elements is easily computed as:

**Table 1.** The Lie bracket operation for $\mathfrak{g} = su(3)$.

| $[\cdot,\cdot]$ | $\mathbf{H_{12}}$ | $\mathbf{H_{23}}$ | $\mathbf{E_{12}}$ | $\mathbf{F_{12}}$ | $\mathbf{E_{23}}$ | $\mathbf{F_{23}}$ | $\mathbf{E_{13}}$ | $\mathbf{F_{13}}$ |
|---|---|---|---|---|---|---|---|---|
| $\mathbf{H_{12}}$ | 0 | 0 | $2F_{12}$ | $-2E_{12}$ | $-F_{23}$ | $E_{23}$ | $F_{13}$ | $-E_{13}$ |
| $\mathbf{H_{23}}$ | * | 0 | $-F_{12}$ | $E_{12}$ | $2F_{23}$ | $-2E_{23}$ | $F_{13}$ | $-E_{13}$ |
| $\mathbf{E_{12}}$ | * | * | 0 | $2H_{12}$ | $E_{13}$ | $F_{13}$ | $-E_{23}$ | $-F_{23}$ |
| $\mathbf{F_{12}}$ | * | * | * | 0 | $F_{13}$ | $-E_{13}$ | $F_{23}$ | $-E_{23}$ |
| $\mathbf{E_{23}}$ | * | * | * | * | 0 | $2H_{23}$ | $E_{12}$ | $F_{12}$ |
| $\mathbf{F_{23}}$ | * | * | * | * | * | 0 | $-F_{12}$ | $E_{12}$ |
| $\mathbf{E_{13}}$ | * | * | * | * | * | * | 0 | $2H_{13}$ |
| $\mathbf{F_{13}}$ | * | * | * | * | * | * | * | 0 |

The *'s below the diagonal remind us that these entries are determined by those above the diagonal, since $[A, B] = -[B, A]$.

By linearity, this table can be used to determine the Lie bracket of *any* pair of elements of $su(3)$. Our clever choice of basis has insured that the table is simple enough for the patterns to become clear. Our goal is to summarize the important patterns in a manner that will generalize to $su(n)$ and beyond.

First, notice that

$$\tau = \text{span}\{H_{12}, H_{23}\} = \{\text{diag}(\lambda_1\mathbf{i}, \lambda_2\mathbf{i}, \lambda_3\mathbf{i}) \mid \lambda_1 + \lambda_2 + \lambda_3 = 0\}$$

is the Lie algebra of the standard maximal torus of $SU(3)$. The other basis elements pair off to span the following three subspaces of $su(3)$, which are called the **root spaces** of $su(3)$:

$$\mathfrak{l}_{12} = \text{span}\{E_{12}, F_{12}\}, \;\; \mathfrak{l}_{23} = \text{span}\{E_{23}, F_{23}\}, \;\; \mathfrak{l}_{13} = \text{span}\{E_{13}, F_{13}\}.$$

We will first study $\mathfrak{l}_{12}$ by asking: how does an element of $\mathfrak{l}_{12}$ bracket with...

(1) ...another element of $\mathfrak{l}_{12}$?

(2) ...an element of $\tau$?

(3) ...an element of a different root space?

More importantly, we will study how these three questions are related, because that's the key to understanding roots.

For question (1), since it is only 2-dimensional, all brackets between pairs of vectors in $\mathfrak{l}_{12}$ can be determined from the single value:

$$\hat{\alpha}_{12} = \left[\frac{E_{12}}{|E_{12}|}, \frac{F_{12}}{|F_{12}|}\right].$$

Table 1 says $[E_{12}, F_{12}] = 2H_{12}$, and therefore $\hat{\alpha}_{12} = H_{12}$. The vector $\hat{\alpha}_{12} \in \tau$ is called a **dual root**. It reports how vectors within its associated root space, $\mathfrak{l}_{12}$, bracket with each other. In $su(3)$ (but not in the generalizations to come), it happens to equal a member of our basis for $\tau$.

For question (2), the relevant block of Table 1 that shows how vectors in $\tau$ bracket with vectors in $\mathfrak{l}_{12}$ is:

| $[\cdot,\cdot]$ | $\mathbf{E_{12}}$ | $\mathbf{F_{12}}$ |
|---|---|---|
| $\mathbf{H_{12}}$ | $2F_{12}$ | $-2E_{12}$ |
| $\mathbf{H_{23}}$ | $-F_{12}$ | $E_{12}$ |

First notice that $\mathfrak{l}_{12}$ is **$\text{ad}_\tau$-invariant**, which means all four entries of this table are members of $\mathfrak{l}_{12}$; that is, $\text{ad}_\tau(\mathfrak{l}_{12}) \subset \mathfrak{l}_{12}$. The specific pattern of these four entries can be summarized as follows: if $X \in \{H_{12}, H_{23}\}$, then

$$[X, E_{12}] = \alpha_{12}(X)F_{12} \text{ and } [X, F_{12}] = -\alpha_{12}(X)E_{12}, \tag{11.1}$$

where

$$\alpha_{12}(H_{12}) = 2, \quad \alpha_{12}(H_{23}) = -1.$$

Let $\alpha_{12} : \tau \to \mathbb{R}$ be the linear function with these values on the basis $\{H_{12}, H_{23}\}$. By linearity, Equation 11.1 is true for all $X \in \tau$. That is, if we identify $\mathfrak{l}_{12} \cong \mathbb{R}^2$ via the ordered basis $\{E_{12}, F_{12}\}$, then for any $X \in \tau$, the linear map $\mathrm{ad}_X : \mathfrak{l}_{12} \to \mathfrak{l}_{12}$ equals a $90°$ counterclockwise rotation composed with a re-scaling by the scaling factor $\alpha_{12}(X)$. The linear function $\alpha_{12} : \tau \to \mathbb{R}$ is called a **root** of $su(3)$. It reports how vectors in $\tau$ bracket with vectors in its associated root space, $\mathfrak{l}_{12}$.

In summary, the dual root $\hat{\alpha}_{12} \in \tau$ encodes the answer to question (1) about $[\mathfrak{l}_{12}, \mathfrak{l}_{12}] \subset \tau$, while the root $\alpha_{12} : \tau \to \mathbb{R}$ encodes the answer to question (2) about $[\tau, \mathfrak{l}_{12}] \subset \mathfrak{l}_{12}$. The key observation is the following relationship between the root and the dual root: for all $X \in \tau$,

$$\alpha_{12}(X) = \langle \hat{\alpha}_{12}, X \rangle_{\mathbb{R}}. \tag{11.2}$$

Recall here that $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ is the inner product on $su(3) \subset M_3(\mathbb{C}) \cong \mathbb{R}^{18}$, as described in Section 4 of Chapter 8 (see in particular Equation 8.7). Equation 11.2 says that the root $\alpha_{12}$ encodes exactly the same information as the dual root $\hat{\alpha}_{12}$; either can be determined from the other, so either encodes the answers to both questions (1) and (2).

Before addressing question (3), we mention that the other two root spaces, $\mathfrak{l}_{23}$ and $\mathfrak{l}_{13}$, work the same way with respect to questions (1) and (2). In particular, for each choice ($ij = 23$ or $ij = 13$) there is an associated dual root $\hat{\alpha}_{ij} \in \tau$ and an associated root $\alpha_{ij} : \tau \to \mathbb{R}$ defined as:

$$\hat{\alpha}_{ij} = \left[ \frac{E_{ij}}{|E_{ij}|}, \frac{F_{ij}}{|F_{ij}|} \right], \quad \alpha_{ij}(X) = \langle \hat{\alpha}_{ij}, X \rangle_{\mathbb{R}} \quad \text{for all } X \in \tau.$$

The root space $\mathfrak{l}_{ij}$ is $\mathrm{ad}_\tau$-invariant, and the manner in which $[\tau, \mathfrak{l}_{ij}] \subset \mathfrak{l}_{ij}$ is governed by:

$$[X, E_{ij}] = \alpha_{ij}(X) F_{ij} \text{ and } [X, F_{ij}] = -\alpha_{ij}(X) E_{ij}.$$

Lastly, we will address question (3) about how elements of one root space bracket with elements of another root space. The relevant blocks of Table 1 are:

| $[\cdot,\cdot]$ | $\mathbf{E_{23}}$ | $\mathbf{F_{23}}$ |
|---|---|---|
| $\mathbf{E_{12}}$ | $E_{13}$ | $F_{13}$ |
| $\mathbf{F_{12}}$ | $F_{13}$ | $-E_{13}$ |

| $[\cdot,\cdot]$ | $\mathbf{E_{13}}$ | $\mathbf{F_{13}}$ |
|---|---|---|
| $\mathbf{E_{12}}$ | $-E_{23}$ | $-F_{23}$ |
| $\mathbf{F_{12}}$ | $F_{23}$ | $-E_{23}$ |

| $[\cdot,\cdot]$ | $\mathbf{E_{13}}$ | $\mathbf{F_{13}}$ |
|---|---|---|
| $\mathbf{E_{23}}$ | $E_{12}$ | $F_{12}$ |
| $\mathbf{F_{23}}$ | $-F_{12}$ | $E_{12}$ |

We see that $[\mathfrak{l}_{12}, \mathfrak{l}_{23}] = \mathfrak{l}_{13}$, $[\mathfrak{l}_{12}, \mathfrak{l}_{13}] = \mathfrak{l}_{23}$ and $[\mathfrak{l}_{23}, \mathfrak{l}_{13}] = \mathfrak{l}_{12}$; that is, each pair of distinct root spaces bracket into the third root space. Except for the signs, the pattern of *how* the basis elements bracket together is exactly the same for all three of the above table-blocks. Assuming we can identify the sign pattern (which we can), this means we need only understand *which* root spaces bracket into which, since the question of *how* they bracket together will always be the same. To foreshadow the general theory, we mention that the answer to this "which" question is secretly encoded in the pattern of which pairs of dual roots have sums/differences that equal $\pm$ another dual root:

$$\begin{aligned} [\mathfrak{l}_{12}, \mathfrak{l}_{23}] = \mathfrak{l}_{13} &\quad\longleftrightarrow\quad \hat{\alpha}_{12} + \hat{\alpha}_{23} = \hat{\alpha}_{13}, \\ [\mathfrak{l}_{12}, \mathfrak{l}_{13}] = \mathfrak{l}_{23} &\quad\longleftrightarrow\quad -\hat{\alpha}_{12} + \hat{\alpha}_{13} = \hat{\alpha}_{23}, \\ [\mathfrak{l}_{23}, \mathfrak{l}_{13}] = \mathfrak{l}_{12} &\quad\longleftrightarrow\quad \hat{\alpha}_{23} - \hat{\alpha}_{13} = -\hat{\alpha}_{12}. \end{aligned}$$

Assuming we can identify the sign pattern (which we can), it seems that the dual roots encode the the full answers to questions (1), (2) and (3).

In summary, each of the three root spaces has a dual root, which is initially defined to encode only the information needed to compute the bracket of any pair of vectors that are both in its corresponding root space. But the dual roots turned out to together encode enough information to determine *all* of Table 1.

## 2. The structure of $\mathfrak{g} = su(n)$

In this section, we discuss how the patterns in $su(3)$ generalize to $su(n)$. Recall that the Lie algebra of the standard maximal torus of $SU(n)$ equals:

$$\tau = \{\text{diag}(\lambda_1 \mathbf{i}, ..., \lambda_n \mathbf{i}) \mid \lambda_1 + \cdots + \lambda_n = 0\}.$$

For each pair $(i, j)$ of distinct integers between 1 and $n$, let $H_{ij} \in \tau$ denote the matrix with $\mathbf{i}$ in position $(i, i)$ and $-\mathbf{i}$ in position $(j, j)$. Let $E_{ij} \in su(n)$ denote the matrix with 1 in position $(i, j)$ and $-1$ in position $(j,i)$. Let $F_{ij} \in su(n)$ denote the matrix with $\mathbf{i}$ in positions $(i,j)$ and $(j,i)$.

Notice $\{H_{12}, H_{23}, ..., H_{(n-1)n}\}$ is a basis for $\tau$. This basis for $\tau$, together with all of the $E$'s and $F$'s for which $i < j$, forms a basis of $su(n)$. In the $n = 3$ case, this is exactly the basis constructed in the previous section. But unlike in the previous section, we are defining $E_{ij}$, $F_{ij}$ and $H_{ij}$ not only when $i < j$ but also when $j < i$. This convention will prove useful even though the added vectors are redundant, since $E_{ji} = -E_{ij}$ and $F_{ji} = F_{ji}$.

The spaces $\mathfrak{l}_{ij} = \text{span}\{E_{ij}, F_{ij}\}$ are called the **root spaces** of $su(n)$. Notice that $\mathfrak{l}_{ij} = \mathfrak{l}_{ji}$, so these are considered two names for the same root space. Notice that the root spaces are orthogonal to $\tau$ and are mutually orthogonal to each other.

For each pair $(i,j)$ of distinct integers between 1 and $n$, define the corresponding **dual root** $\hat{\alpha}_{ij} \in \tau$ and **root** $\alpha_{ij} : \tau \to \mathbb{R}$ exactly as before:

$$\hat{\alpha}_{ij} = \left[\frac{E_{ij}}{|E_{ij}|}, \frac{F_{ij}}{|F_{ij}|}\right], \quad \alpha_{ij}(X) = \langle \hat{\alpha}_{ij}, X \rangle_{\mathbb{R}} \quad \text{for all } X \in \tau.$$

As before, each root space, $\mathfrak{l}_{ij}$, is $\text{ad}_\tau$-invariant, and the manner in which $[\tau, \mathfrak{l}_{ij}] \subset \mathfrak{l}_{ij}$ is governed by:

$$[X, E_{ij}] = \alpha_{ij}(X) F_{ij} \text{ and } [X, F_{ij}] = -\alpha_{ij}(X) E_{ij}.$$

This says that if we identify $\mathfrak{l}_{ij} \cong \mathbb{R}^2$ via the ordered basis $\{E_{ij}, F_{ij}\}$, then for any $X \in \tau$, the linear map $\text{ad}_X : \mathfrak{l}_{ij} \to \mathfrak{l}_{ij}$ equals a $90°$ counterclockwise rotation composed with a re-scaling by the scaling factor $\alpha_{ij}(X)$.

If $i < j$, then $\alpha_{ij}$ is called a **positive root** and $\hat{\alpha}_{ij}$ is called a **positive dual root**; these were the only type considered in the previous section. Notice that $\hat{\alpha}_{ji} = -\hat{\alpha}_{ij}$ and $\alpha_{ji} = -\alpha_{ij}$, so the remaining roots are redundant to the positive ones, which is why we only considered positive roots in the previous section.

In the case $n = 3$ from the previous section, there were exactly three root spaces, and each pair of them bracketed into the third. But when $n > 3$, there are pairs of root spaces of $su(n)$ that bracket to zero. For example, check that $[\mathfrak{l}_{12}, \mathfrak{l}_{34}] = 0$ in $su(4)$. The reason is that there are no common indices; that is, the sets $\{1,2\}$ and $\{3,4\}$ are disjoint.

Here is the most elegant way to completely describe the bracket between any pair of distinct root spaces in $su(n)$. If they have no common index, then they bracket to zero. Otherwise, their bracket structure is described by the following general rule: $[\mathfrak{l}_{ij}, \mathfrak{l}_{jk}] = \mathfrak{l}_{ik}$, with brackets of individual basis elements given by:

| $[\cdot,\cdot]$ | $\mathbf{E_{jk}}$ | $\mathbf{F_{jk}}$ |
|---|---|---|
| $\mathbf{E_{ij}}$ | $E_{ik}$ | $F_{ik}$ |
| $\mathbf{F_{ij}}$ | $F_{ik}$ | $-E_{ik}$ |

This table is arranged with the second index of the first $\mathfrak{l}$ equalling the first index of the second. In the previous section, we were not always able to respect this index-ordering convention because we had only defined $\mathfrak{l}_{ij}$, $F_{ij}$ and $E_{ij}$ when $i < j$. Nevertheless, this table can be converted into the tables in the previous section for $[\mathfrak{l}_{12}, \mathfrak{l}_{13}] \subset \mathfrak{l}_{23}$ and $[\mathfrak{l}_{23}, \mathfrak{l}_{13}] \subset \mathfrak{l}_{12}$ simply by using that $\mathfrak{l}_{ji} = \mathfrak{l}_{ij}$, $E_{ji} = -E_{ij}$ and $F_{ji} = F_{ij}$ as needed. In particular, the signs in the tables from the previous section are accounted for by this general rule.

As before, the dual roots secretly encode this information about which pairs of root spaces bracket into which. The general relationship is:

$$[\mathfrak{l}_{ij}, \mathfrak{l}_{jk}] = \mathfrak{l}_{ik} \quad \longleftrightarrow \quad \hat{\alpha}_{ij} + \hat{\alpha}_{jk} = \hat{\alpha}_{ik}. \tag{11.3}$$

The corresponding relationships from the previous section all come from this general one (together with the rule that $\hat{\alpha}_{ji} = -\hat{\alpha}_{ij}$). The signs appeared in the previous section because we were working only with *positive* dual roots, which left us unable to respect the index ordering convention of Equation 11.3.

Furthermore, the fact that $[\mathfrak{l}_{12}, \mathfrak{l}_{34}] = 0$ in $su(4)$ converts into the fact that $\hat{\alpha}_{12} \pm \hat{\alpha}_{34}$ is not a dual root. Here is a description, purely in terms of dual roots, of which root spaces bracket into which. If either $\hat{\alpha}_{ab} + \hat{\alpha}_{cd}$ or $\hat{\alpha}_{ab} - \hat{\alpha}_{cd}$ equals a dual root, $\hat{\alpha}_{ij}$ (they never both equal a dual root), then $[\mathfrak{l}_{ab}, \mathfrak{l}_{cd}] = \mathfrak{l}_{ij}$. Otherwise, $[\mathfrak{l}_{ab}, \mathfrak{l}_{cd}] = 0$.

In summary, each root space of $su(n)$ has two dual roots (which are redundant because they are the negatives of each other). Each dual root was initially defined to encode only the information needed to compute the bracket of any pair of vectors that are both within its corresponding root space. But the dual roots turned out to together encode enough information to determine the *entire* Lie bracket structure of $su(n)$.

## 3. An invariant decomposition of $\mathfrak{g}$

In this and the next two sections, we generalize to an arbitrary compact Lie group all of the structures and patterns that we previously observed for $SU(n)$.

Let $G$ be a compact path-connected Lie group with Lie algebra $\mathfrak{g}$. Let $T \subset G$ be a maximal torus, with Lie algebra $\tau \subset \mathfrak{g}$. Let $\langle \cdot, \cdot \rangle$ be an Ad-invariant inner product on $\mathfrak{g}$, as guaranteed by Theorem 10.32. Recall that "Ad-invariant" means $\langle \mathrm{Ad}_g X, \mathrm{Ad}_g Y \rangle = \langle X, Y \rangle$ for all $g \in G$ and all $X, Y \in \mathfrak{g}$. As in Proposition 8.14, this implies "infinitesimal Ad-invariance": for all $A, B, C \in \mathfrak{g}$,

$$\langle [A, B], C \rangle = -\langle [A, C], B \rangle.$$

In addition to these equations, we will repeatedly use the facts about maximal tori summarized in Proposition 10.29.

We now begin to generalize to $G$ the patterns we have observed for $SU(n)$, beginning with:

**Theorem 11.1.** $\mathfrak{g}$ *decomposes as an orthogonal direct sum,*

$$\mathfrak{g} = \tau \oplus \mathfrak{l}_1 \oplus \mathfrak{l}_2 \oplus \cdots \oplus \mathfrak{l}_m,$$

*where each* $\mathfrak{l}_i$ *is a 2-dimensional* $Ad_T$*-invariant subspace of* $\mathfrak{g}$.

The spaces $\{\mathfrak{l}_i\}$ are called the **root spaces** of $G$. Each root space is $\mathrm{Ad}_T$-invariant, which means that for each $g \in T$ and each $V \in \mathfrak{l}_i$, we have $\mathrm{Ad}_g(V) \in \mathfrak{l}_i$. In other words, $\mathrm{Ad}_T(\mathfrak{l}_i) \subset \mathfrak{l}_i$. By the definition of the Lie bracket, this implies that each root space $\mathfrak{l}_i$ is also $\mathrm{ad}_\tau$-invariant, which means that $\mathrm{ad}_\tau(\mathfrak{l}_i) \subset \mathfrak{l}_i$.

Since $2m = \dim(G) - \mathrm{rank}(G)$, the theorem implies that $\dim(G) - \mathrm{rank}(G)$ is even.

**Proof.** For each $g \in T$, the linear function $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ restricts to $\tau$ as the identity function, because $T$ is abelian. Therefore, $\mathrm{Ad}_g$ sends any vector $V \in \tau^\perp$ to another vector in $\tau^\perp$, since for all $X \in \tau$,

$$\langle \mathrm{Ad}_g V, X \rangle = \langle V, \mathrm{Ad}_{g^{-1}} X \rangle = \langle V, X \rangle = 0.$$

We choose a fixed orthonormal basis, $\mathcal{B}$, of $\tau^\perp$, via which we identify $\tau^\perp \cong \mathbb{R}^s$, where $s = \dim(\tau^\perp) = \dim(G) - \mathrm{rank}(G)$. For each $g \in T$, the map $\mathrm{Ad}_g : \tau^\perp \to \tau^\perp$ can be represented with respect to $\mathcal{B}$ as left multiplication by some matrix in $O(s)$; in this way, we can consider Ad as a smooth homomorphism $\mathrm{Ad} : T \to O(s)$.

Since $T$ is a compact abelian path-connected Lie group, so must be its image under a smooth homomorphism. Theorem 9.5 generalizes to say that any compact abelian path-connected Lie group is isomorphic to a torus. Thus, the image, $\mathrm{Ad}(T) \subset O(s)$, is a torus in $O(s)$. Let $\tilde{T}$ be a maximal torus of $O(s)$ that contains $\mathrm{Ad}(T)$. By Proposition 10.29(2), $\tilde{T}$ equals a conjugate of the standard maximal torus of $O(s)$. Said differently, after conjugating our basis $\mathcal{B}$, we can assume that $\tilde{T}$ equals the standard maximal torus of $O(s)$.

If $s$ is even, so that $s = 2m$ for some integer $m$, then we'll write this newly conjugated orthonormal basis of $\tau^\perp$ as

$$\mathcal{B} = \{E_1, F_1, E_2, F_2, ..., E_m, F_m\}.$$

Our description in Chapter 9 of the standard maximal torus of $O(2m)$ shows that each of the spaces $\mathfrak{l}_i = \mathrm{span}\{E_i, F_i\}$ is invariant under the left-multiplication map, $L_a$, for all $a \in \tilde{T}$. This means that $L_a(\mathfrak{l}_i) \subset \mathfrak{l}_i$. In particular, this holds for all $a \in \tilde{T}$ for which $L_a$ represents $\mathrm{Ad}_g$ for some $g \in T$. Therefore, each $\mathfrak{l}_i$ is $Ad_T$-invariant.

It remains to demonstrate that $s$ cannot be odd. If $s$ were odd, so that $s = 2m+1$, then one more element, $V \in \tau^\perp$, would need to be added to the above basis $\mathcal{B}$. It follows from our description in Chapter 9 of the standard maximal torus of $O(2m+1)$ that $L_a(V) = V$ for all $a \in \tilde{T}$, so $Ad_g(V) = V$ for all $g \in T$. Therefore, $[X, V] = 0$ for all $X \in \tau$, contradicting Proposition 10.29 (4). $\square$

## 4. The definition of roots and dual roots

Decompose $\mathfrak{g} = \tau \oplus \mathfrak{l}_1 \oplus \cdots \oplus \mathfrak{l}_m$, as in Theorem 11.1. For each $i$, let $\{E_i, F_i\}$ be an orthonormal ordered basis for $\mathfrak{l}_i$.

**Definition 11.2.** *For each $i$, define $\hat{\alpha}_i = [E_i, F_i]$, and define the linear function $\alpha_i : \tau \to \mathbb{R}$ such that for all $X \in \tau$, $\alpha_i(X) = \langle \hat{\alpha}_i, X \rangle$.*

Notice that $\alpha_i$ and $\hat{\alpha}_i$ contain the same information. The next proposition shows that the $\alpha$'s determine how vectors in $\tau$ bracket with vectors in the root spaces.

**Proposition 11.3.** *For each $i$, $\hat{\alpha}_i \in \tau$. Further, for all $X \in \tau$,*

$$[X, E_i] = \alpha_i(X) \cdot F_i \text{ and } [X, F_i] = -\alpha_i(X) \cdot E_i.$$

**Proof.** Let $X \in \tau$. Since $\mathfrak{l}_i$ is $\mathrm{ad}_\tau$-invariant, we know $[X, E_i] \in \mathfrak{l}_i$. Also, $[X, E_i]$ is orthogonal to $\mathfrak{l}_i$'s first basis vector, $E_i$, since

$$\langle [X, E_i], E_i \rangle = -\langle [E_i, X], E_i \rangle = \langle [E_i, E_i], X \rangle = 0,$$

so $[X, E_i]$ must be a multiple of $\mathfrak{l}_i$'s second basis vector, $F_i$. That is, $[X, E_i] = \lambda \cdot F_i$. This multiple is:

$$\lambda = \langle [X, E_i], F_i \rangle = -\langle [E_i, X], F_i \rangle = \langle [E_i, F_i], X \rangle = \langle \hat{\alpha}_i, X \rangle = \alpha_i(X).$$

Similarly, $[X, F_i] = -\alpha_i(X) \cdot E_i$.

Finally, we prove that $\hat{\alpha}_i = [E_i, F_i] \in \tau$. Using the Jacobi identity, we have for all $X \in \tau$ that:

$$\begin{aligned} [X, [E_i, F_i]] &= [E_i, [X, F_i]] - [F_i, [X, E_i]] \\ &= -[E_i, \alpha_i(X) \cdot E_i] - [F_i, \alpha(X) \cdot F_i] = 0. \end{aligned}$$

Since $[E_i, F_i]$ commutes with every $X \in \tau$, we know $[E_i, F_i] \in \tau$. $\square$

For each $i$, the linear function $\alpha_i : \tau \to \mathbb{R}$ records the initial speed at which each vector $X \in \tau$ rotates the root space $\mathfrak{l}_i$. To understand this remark, notice that the function $\mathrm{ad}_X : \mathfrak{l}_i \to \mathfrak{l}_i$ (which sends $A \mapsto [X, A]$) is given with respect to the ordered basis $\{E_i, F_i\}$ as left multiplication by the "infinitesimal rotation matrix":

$$\mathrm{ad}_X = \begin{pmatrix} 0 & -\alpha_i(X) \\ \alpha_i(X) & 0 \end{pmatrix}.$$

Thus, the function $Ad_{e^{tX}} : \mathfrak{l}_i \to \mathfrak{l}_i$ is given in this ordered basis as left multiplication by the rotation matrix:

$$\mathrm{Ad}_{e^{tX}} = e^{\mathrm{ad}_X} = \begin{pmatrix} \cos(\alpha_i(X)t) & -\sin(\alpha_i(X)t) \\ \sin(\alpha_i(X)t) & \cos(\alpha_i(X)t) \end{pmatrix}.$$

For each index $i$, let $R_i : \mathfrak{l}_i \to \mathfrak{l}_i$ denote a $90^\circ$ rotation of $\mathfrak{l}_i$ that is "counterclockwise" with respect to the ordered basis $\{E_i, F_i\}$. That is, $R_i$ is the linear function for which $R_i(E_i) = F_i$ and $R_i(F_i) = -E_i$. Notice that for all $V \in \mathfrak{l}_i$ and all $X \in \tau$, we have:

$$\begin{aligned} \mathrm{ad}_X(V) &= \alpha_i(X) \cdot R_i(V), \\ \mathrm{Ad}_{e^{tX}} V &= \cos(\alpha_i(X)t) \cdot V + \sin(\alpha_i(X)t) \cdot R_i(V), \end{aligned}$$

simply because these equations are true on a basis $V \in \{E_i, F_i\}$. More generally, since $\mathfrak{g} = \tau \oplus \mathfrak{l}_1 \oplus \cdots \oplus \mathfrak{l}_m$ is an orthogonal decomposition, each $V \in \mathfrak{g}$ decomposes uniquely as $V = V^0 + V^1 + \cdots + V^m$, with $V^0 \in \tau$ and with $V^i \in \mathfrak{l}_i$ for each $1 \leq i \leq m$. For each $X \in \tau$, $\mathrm{ad}_X$ and $\mathrm{Ad}_{e^{tX}}$ act independently on the $\mathfrak{l}$'s, so:

$$\mathrm{ad}_X V = \sum_{i=1}^{m} \alpha_i(X) \cdot R_i(V^i), \tag{11.4}$$

$$\mathrm{Ad}_{e^{tX}} V = V^0 + \sum_{i=1}^{m} \cos(\alpha_i(X)t) \cdot V^i + \sin(\alpha_i(X)t) \cdot R_i(V^i). \tag{11.5}$$

Thus, the one-parameter group $t \mapsto \mathrm{Ad}_{e^{tX}}$ independently rotates each $\mathfrak{l}_i$ with period $2\pi/|\alpha_i(X)|$. The rotation is counterclockwise if $\alpha_i(X) > 0$ and clockwise if $\alpha_i(X) < 0$.

We caution that there is generally no basis-independent notion of clockwise. If we replace the ordered basis $\{E_i, F_i\}$ with $\{E_i, -F_i\}$ (or with $\{F_i, E_i\}$), this causes $R_i$ and $\alpha_i$ to each be multiplied by $-1$, so our notion of clockwise is reversed. Nevertheless, this sign ambiguity is the only sense in which our definition of $\alpha_i$ is basis-dependent. The absolute value (or equivalently the square) of each $\alpha_i$ is basis-independent:

**Proposition 11.4.** *Each function $\alpha_i^2 : \tau \to \mathbb{R}^{\geq 0}$ is independent of the choice of ordered orthonormal basis $\{E_i, F_i\}$ for $\mathfrak{l}_i$.*

**Proof.** Let $X \in \tau$. Consider the linear function $\mathrm{ad}_X^2 : \mathfrak{l}_i \to \mathfrak{l}_i$, which sends $V \to \mathrm{ad}_X(\mathrm{ad}_X(V)) = [X,[X,V]]$. By Proposition 11.3, we have $\mathrm{ad}_X^2(E_i) = -\alpha_i(X)^2 \cdot E_i$ and $\mathrm{ad}_X^2(F_i) = -\alpha_i(X)^2 \cdot F_i$. Thus,

$$\mathrm{ad}_X^2 = -\alpha_i(X)^2 \cdot \mathrm{Id}.$$

Therefore $-\alpha_i(X)^2$ is an eigenvalue of $\mathrm{ad}_X^2$ and so is basis-independent. □

For our general definition, we will use:

**Definition 11.5.** *A non-zero linear function $\alpha : \tau \to \mathbb{R}$ is called a **root** of $G$ if there exists a 2-dimensional subspace $\mathfrak{l} \subset \mathfrak{g}$ with an ordered orthonormal basis $\{E, F\}$ such that for each $X \in \tau$, we have*

$$[X, E] = \alpha(X) \cdot F \quad \textit{and} \quad [X, F] = -\alpha(X) \cdot E.$$

*In this case, $\mathfrak{l}$ is called the **root space** for $\alpha$, and the **dual root** for $\alpha$ means the unique vector $\hat{\alpha} \in \tau$ such that $\alpha(X) = \langle \hat{\alpha}, X \rangle$ for all $X \in \tau$.*

The fact that a unique such dual root vector always exists is justified in Exercise 11.2.

Notice that if $\alpha$ is a root of $G$ with root space $\mathfrak{l} = \mathrm{span}\{E, F\}$, then $-\alpha$ is a root of $G$ with the same root space $\mathfrak{l} = \mathrm{span}\{F, E\}$.

**Proposition 11.6.** *The functions $\pm\alpha_1, ..., \pm\alpha_m$ from Definition 11.2 are all roots of $G$.*

**Proof.** It only remains to show that each $\alpha_i$ is non-zero (not the zero function). But if $\alpha_i(X) = 0$ for all $X \in \tau$, then $E_i$ and $F_i$ would commute with every element of $\tau$, contradicting Proposition 10.29 (4). □

Typically, $m > \dim(\tau)$, so the set $\{\hat{\alpha}_1, ..., \hat{\alpha}_m\}$ is too big to be a basis of $\tau$, but we at least have:

**Proposition 11.7.** *If the center of $G$ is finite, then the dual roots of $G$ span $\tau$.*

**Proof.** If some $X \in \tau$ were orthogonal to all of the dual roots, then $[X, A] = 0$ for all $A \in \mathfrak{g}$, and therefore $e^{tA}$ would lie in the center of $G$ for all $t \in \mathbb{R}$. □

Recall that $SO(n)$ (when $n > 2$), $SU(n)$ and $Sp(n)$ have finite centers according to Theorem 9.10. Even though the dual roots are typically linearly dependent, we at least have the following result, whose proof is beyond the scope of this book:

**Lemma 11.8.** *No pair of the dual roots $\{\hat{\alpha}_1, ..., \hat{\alpha}_m\}$ is equal (or even parallel) to each other.*

We require this lemma to prove that $\{\pm\alpha_1, ..., \pm\alpha_m\}$ are the only roots. You will observe in Exercise 11.19 that if the lemma were false, then there would be other roots.

**Definition 11.9.** *A vector $X \in \tau$ is called a **regular vector** if the following are distinct non-zero numbers: $\alpha_1(X)^2, ..., \alpha_m(X)^2$.*

For example, when $G = SU(n)$, $X = \text{diag}(\lambda_1 \mathbf{i}, ..., \lambda_n \mathbf{i})$ is regular if and only if no difference of two $\lambda$'s equals zero or equals the difference of another two $\lambda$'s.

**Proposition 11.10.** *The regular vectors of $G$ form an open dense subset of $\tau$. In particular, regular vectors exist.*

**Proof.** Exercise 11.14, using Lemma 11.8. $\square$

**Proposition 11.11.** *If $X \in \tau$ is regular, then the map $ad_x^2 : \mathfrak{g} \to \mathfrak{g}$ has eigenvalues $0, -\alpha_1(X)^2, -\alpha_2(X)^2, ..., -\alpha_m(X)^2$ with corresponding eigenspaces $\tau, \mathfrak{l}_1, \mathfrak{l}_2, ..., \mathfrak{l}_m$.*

This proposition follows from the proof of Proposition 11.4. It says that if $X$ is regular, then the decomposition of $\mathfrak{g}$ into eigenspaces of $\text{ad}_X^2$ is the same as the root space decomposition of $\mathfrak{g}$ from Theorem 11.1, with $\tau$ equal to the kernel of $\text{ad}_X^2$.

If $X \in \tau$ is not regular, then the eigenspace decomposition of $\text{ad}_X^2$ is "coarser" than the regular one; that is, the root spaces $\mathfrak{l}_i$ for which $\alpha_i(X) = 0$ are grouped with $\tau$ to form the kernel of $\text{ad}_X^2$, and each other eigenspace is a root space or a sum of root spaces, $\mathfrak{l}_{i_1} \oplus \cdots \oplus \mathfrak{l}_{i_k}$, coming from repeated values $\alpha_{i_1}(X)^2 = \cdots = \alpha_{i_k}(X)^2$. Since $\mathfrak{g}$'s decomposition into root spaces corresponds to the "finest" of the $\text{ad}_X^2$ eigenspace decompositions, this decomposition is unique. We have just established:

**Proposition 11.12.** *The decomposition from Theorem* 11.1 *is unique, and therefore* $\{\pm\alpha_1, ..., \pm\alpha_m\}$ *are the only roots of* $G$*.*

## 5. The bracket of two root spaces

The roots describe exactly how vectors in $\tau$ bracket with vectors in the root spaces. Surprisingly, they also help determine how vectors in one root space bracket with vectors in another root space. To explain how, we establish the following notation. If $\hat{\alpha}_i + \hat{\alpha}_j$ equals a dual root, then let $\mathfrak{l}_{ij}^+$ denote its root space; otherwise, let $\mathfrak{l}_{ij}^+ = \{0\}$. If $\hat{\alpha}_i - \hat{\alpha}_j$ equals a dual root, then let $\mathfrak{l}_{ij}^-$ denote its root space; otherwise, let $\mathfrak{l}_{ij}^- = \{0\}$. With this notation:

**Theorem 11.13.** $[\mathfrak{l}_i, \mathfrak{l}_j] \subset \mathfrak{l}_{ij}^+ \oplus \mathfrak{l}_{ij}^-$.

In particular, if neither $\hat{\alpha}_i + \hat{\alpha}_j$ nor $\hat{\alpha}_i - \hat{\alpha}_j$ equals a dual root, then $[\mathfrak{l}_i, \mathfrak{l}_j] = \{0\}$. For all of the classical groups except $SO(2n+1)$, we'll see that the sum and difference never both equal dual roots, so any pair of root spaces must bracket to zero or to a single root space:

**Corollary 11.14.** *If* $G \in \{SU(n), SO(2n), Sp(n)\}$*, then for any pair* $(i, j)$*, either* $[\mathfrak{l}_i, \mathfrak{l}_j] = 0$ *or there exists* $k$ *such that* $[\mathfrak{l}_i, \mathfrak{l}_j] \subset \mathfrak{l}_k$*. In the latter case,* $\hat{\alpha}_i \pm \hat{\alpha}_j = \pm\hat{\alpha}_k$*.*

Section 11 will contain the standard proof of Theorem 11.13 using complexified Lie algebras. For now, we offer the following longer but less abstract proof:

**Proof of Theorem 11.13.** Let $V \in \mathfrak{l}_i$ and let $W \in \mathfrak{l}_j$, and define $U = [V, W]$. We wish to prove that $U \in \mathfrak{l}_{ij}^+ \oplus \mathfrak{l}_{ij}^-$. First notice that $U \in \tau^\perp$ because for all $Y \in \tau$,

$$\langle U, Y\rangle = \langle [V, W], Y\rangle = -\langle [V, Y], W\rangle = \langle [Y, V], W\rangle = 0.$$

For any $X \in \tau$, we can define the path $U_t = [V_t, W_t]$, where

$$V_t = \mathrm{Ad}_{e^{tX}} V = \cos(\alpha_i(X)t) \cdot V + \sin(\alpha_i(X)t) \cdot R_i(V),$$
$$W_t = \mathrm{Ad}_{e^{tX}} W = \cos(\alpha_j(X)t) \cdot W + \sin(\alpha_j(X)t) \cdot R_j(W)$$

are circles in $\mathfrak{l}_i$ and $\mathfrak{l}_j$. Standard trigonometric identities yield:

$$\begin{aligned} 2U_t \quad = \quad & \cos((\alpha_i(X)+\alpha_j(X))t)([V,W]-[R_iV,R_jW]) \\ & +\sin((\alpha_i(X)+\alpha_j(X))t)([R_iV,W]+[V,R_jW]) \\ & +\cos((\alpha_i(X)-\alpha_j(X))t)([V,W]+[R_iV,R_jW]) \\ & +\sin((\alpha_i(X)-\alpha_j(X))t)([R_iV,W]-[V,R_jW]). \end{aligned} \tag{11.6}$$

On the other hand, since

$$U_t = [\mathrm{Ad}_{e^{tX}}V, \mathrm{Ad}_{e^{tX}}W] = \mathrm{Ad}_{e^{tX}}[V,W] = \mathrm{Ad}_{e^{tX}}U,$$

we can decompose $U = U^1 + \cdots + U^m$ (with $U^i \in \mathfrak{l}_i$), and Equation 11.5 gives:

$$U_t = \sum_{k=1}^{m} \cos(\alpha_k(X)t)\cdot U^k + \sin(\alpha_k(X)t)\cdot R_k(U^k). \tag{11.7}$$

The $k^{\text{th}}$ term of this sum, denoted $U_t^k$, is a circle in $\mathfrak{l}_k$.

For any $X \in \tau$, the expressions for $U_t$ obtained from Equations 11.6 and 11.7 must equal each other. We claim this implies that the first two lines of Equation 11.6 must form a circle in $\mathfrak{l}_{ij}^+$ and the last two lines must form a circle in $\mathfrak{l}_{ij}^-$, so in particular $U_t \in \mathfrak{l}_{ij}^+ \oplus \mathfrak{l}_{ij}^-$ as desired. This implication is perhaps most easily seen by considering special types of vectors $X \in \tau$, as follows.

First, choose $X \perp \mathrm{span}\{\hat{\alpha}_i, \hat{\alpha}_j\}$, so $\alpha_i(X) = \alpha_j(X) = 0$, so $t \mapsto U_t$ is constant. If some $U^k \neq 0$, then $\alpha_k(X) = 0$, which means that $\hat{\alpha}_k \perp X$. In summary, if $U^k \neq 0$, then $\hat{\alpha}_k$ must be perpendicular to any $X$ which is perpendicular to $\mathrm{span}\{\hat{\alpha}_i, \hat{\alpha}_j\}$. We conclude that if $U^k \neq 0$, then $\hat{\alpha}_k \in \mathrm{span}\{\hat{\alpha}_i, \hat{\alpha}_j\}$.

Next, choose $X \in \mathrm{span}\{\hat{\alpha}_i, \hat{\alpha}_j\}$ with $\alpha_i(X) = \langle \hat{\alpha}_i, X\rangle = 1$ and $\alpha_j(X) = \langle \hat{\alpha}_j, X\rangle = 0$, which is possible because $\hat{\alpha}_i$ and $\hat{\alpha}_j$ are not parallel, by Lemma 11.8. If some $U^k \neq 0$, then Equation 11.6 shows that $t \mapsto U_t^k$ has period $= 2\pi$, so $\alpha_k(X) = \langle \hat{\alpha}_k, X\rangle = \pm 1$, by Equation 11.7. In summary, if some $U^k \neq 0$, then $\hat{\alpha}_k \in \mathrm{span}\{\hat{\alpha}_i, \hat{\alpha}_j\}$ has the same projection onto the orthogonal complement of $\hat{\alpha}_j$ as does $\pm\hat{\alpha}_i$. Reversing the roles of $i$ and $j$ shows that $\hat{\alpha}_k$ also has the same projection onto the orthogonal complement of $\hat{\alpha}_i$ as does $\pm\hat{\alpha}_j$. It follows easily that $\hat{\alpha}_k = \pm\hat{\alpha}_i \pm \hat{\alpha}_j$, so $U \in \mathfrak{l}_{ij}^+ \oplus \mathfrak{l}_{ij}^-$. $\square$

For a single $V \in \mathfrak{l}_i$ and $W \in \mathfrak{l}_j$, suppose we know the bracket $[V, W] = A^+ + A^-$ (with $A^+ \in \mathfrak{l}_{ij}^+$ and $A^- \in \mathfrak{l}_{ij}^-$). This single bracket determines the entire bracket operation between $\mathfrak{l}_i$ and $\mathfrak{l}_j$. To see how, let $R_+$ denote the $90°$ rotation of $\mathfrak{l}_{ij}^+$ that is counterclockwise if $\hat{\alpha}_i + \hat{\alpha}_j = \hat{\alpha}_k$ or clockwise if $\hat{\alpha}_i + \hat{\alpha}_j = -\hat{\alpha}_k$ for some $k$. Similarly let $R_-$ denote the $90°$ rotation of $\mathfrak{l}_{ij}^-$ that is counterclockwise if $\hat{\alpha}_i - \hat{\alpha}_j = \hat{\alpha}_k$ or clockwise if $\hat{\alpha}_i - \hat{\alpha}_j = -\hat{\alpha}_k$. With this notation, the ideas of the previous proof yield the following generalization of our previous table:

**Table 2.** The bracket $[\mathfrak{l}_i, \mathfrak{l}_j] \subset \mathfrak{l}_{ij}^+ \oplus \mathfrak{l}_{ij}^-$.

| $[\cdot,\cdot]$ | **W** | $\mathbf{R_j W}$ |
|---|---|---|
| **V** | $A^+ + A^-$ | $R_+(A^+) - R_-(A^-)$ |
| $\mathbf{R_i V}$ | $R_+(A^+) - R_-(A^-)$ | $-A^+ + A^-$ |

## 6. The structure of $so(2n)$

Let $n > 1$ and $G = SO(2n)$, so $\mathfrak{g} = so(2n)$. Recall that the Lie algebra of the standard maximal torus of $G$ is:

$$\tau = \left\{ \mathrm{diag}\left( \begin{pmatrix} 0 & \theta_1 \\ -\theta_1 & 0 \end{pmatrix}, ..., \begin{pmatrix} 0 & \theta_n \\ -\theta_n & 0 \end{pmatrix} \right) \mid \theta_i \in \mathbb{R} \right\}.$$

Let $H_i \in \tau$ denote the matrix with $\theta_i = 1$ and all other $\theta$'s zero, so that $\{H_1, ..., H_n\}$ is a basis for $\tau$. Also define:

$$E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, F = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, Y = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Think of a matrix in $so(2n)$ as being an $n \times n$ grid of $2 \times 2$ blocks. For each pair $(i, j)$ of distinct indices between 1 and $n$, define $E_{ij}$ so that its $(i, j)^{\text{th}}$ block equals $E$ and its $(j, i)^{\text{th}}$ block equals -$E^T$ and all other blocks are zero. Similarly define $F_{ij}$, $X_{ij}$ and $Y_{ij}$. A basis of $\mathfrak{g}$ is formed from $\{H_1, ..., H_n\}$ together with all $E$'s, $F$'s, $X$'s and $Y$'s with $i < j$. In the case $n = 2$, this basis looks like:

$$H_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad E_{12} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad X_{12} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix},$$

$$H_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad F_{12} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \quad Y_{12} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Define $\mathfrak{l}_{ij} = \text{span}\{E_{ij}, F_{ij}\}$ and $\mathfrak{k}_{ij} = \text{span}\{X_{ij}, Y_{ij}\}$. The decomposition of $\mathfrak{g}$ into root spaces is:

$$\mathfrak{g} = \tau \oplus \{\mathfrak{l}_{ij} \mid i < j\} \oplus \{\mathfrak{k}_{ij} \mid i < j\}.$$

Since $[E_{ij}, F_{ij}] = 2(H_i - H_j)$ and $[X_{ij}, Y_{ij}] = 2(H_i + H_j)$, the dual roots are the following matrices (and their negatives):

$$\hat{\alpha}_{ij} = \left[\frac{E_{ij}}{|E_{ij}|}, \frac{F_{ij}}{|F_{ij}|}\right] = \frac{1}{2}(H_i - H_j),$$
$$\hat{\beta}_{ij} = \left[\frac{X_{ij}}{|X_{ij}|}, \frac{Y_{ij}}{|Y_{ij}|}\right] = \frac{1}{2}(H_i + H_j).$$

The following lists all sums and differences of dual roots that equal dual roots. In each case, a sample bracket value is provided:

$$\begin{array}{lll} \hat{\alpha}_{ij} + \hat{\alpha}_{jk} = \hat{\alpha}_{ik} & [\mathfrak{l}_{ij}, \mathfrak{l}_{jk}] \subset \mathfrak{l}_{ik} & [E_{ij}, E_{jk}] = E_{ik} \\ \hat{\beta}_{ij} - \hat{\beta}_{jk} = \hat{\alpha}_{ik} & [\mathfrak{k}_{ij}, \mathfrak{k}_{jk}] \subset \mathfrak{l}_{ik} & [X_{ij}, X_{jk}] = E_{ik} \\ \hat{\alpha}_{ij} + \hat{\beta}_{jk} = \hat{\beta}_{ik} & [\mathfrak{l}_{ij}, \mathfrak{k}_{jk}] \subset \mathfrak{k}_{ik} & [E_{ij}, X_{jk}] = X_{ik}. \end{array}$$

The brackets of any pair of basis elements can be determined from the above sample bracket values via Table 2, yielding:

**Table 3.** The non-zero bracket relations for $\mathfrak{g} = so(2n)$.

| $[\cdot,\cdot]$ | $\mathbf{E_{jk}}$ | $\mathbf{F_{jk}}$ |
|---|---|---|
| $\mathbf{E_{ij}}$ | $E_{ik}$ | $F_{ik}$ |
| $\mathbf{F_{ij}}$ | $F_{ik}$ | $-E_{ik}$ |

| $[\cdot,\cdot]$ | $\mathbf{X_{jk}}$ | $\mathbf{Y_{jk}}$ |
|---|---|---|
| $\mathbf{X_{ij}}$ | $E_{ik}$ | $-F_{ik}$ |
| $\mathbf{Y_{ij}}$ | $F_{ik}$ | $E_{ik}$ |

| $[\cdot,\cdot]$ | $\mathbf{X_{jk}}$ | $\mathbf{Y_{jk}}$ |
|---|---|---|
| $\mathbf{E_{ij}}$ | $X_{ik}$ | $Y_{ik}$ |
| $\mathbf{F_{ij}}$ | $Y_{ik}$ | $-X_{ik}$ |

## 7. The structure of $so(2n+1)$

Let $n > 0$ and $G = SO(2n+1)$, so $\mathfrak{g} = so(2n+1)$. Recall that the Lie algebra of the standard maximal torus of $G$ is:

$$\tau = \left\{ \mathrm{diag}\left( \begin{pmatrix} 0 & \theta_1 \\ -\theta_1 & 0 \end{pmatrix}, ..., \begin{pmatrix} 0 & \theta_n \\ -\theta_n & 0 \end{pmatrix}, 0 \right) \mid \theta_i \in \mathbb{R} \right\}.$$

Each of the previously defined elements of $so(2n)$ can be considered as an element of $so(2n+1)$ simply by adding a final row and final column of zeros. In order to complete our previous basis of $so(2n)$ to a basis of $so(2n+1)$, we need the following additional matrices. For each $1 \leq i \leq n$, let $W_i \in so(2n+1)$ denote the matrix with entry $(2i-1, 2n+1)$ equal to 1 and entry $(2n+1, 2i-1)$ equal to $-1$, and all other entries equal to zero. Let $V_i$ denote the matrix with entry $(2i, 2n+1)$ equal to 1 and entry $(2n+1, 2i)$ equal to $-1$, and all other entries equal to zero. For $n = 2$, these extra basis elements are:

$$W_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad V_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \end{pmatrix},$$

$$W_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix}, \qquad V_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix}.$$

Define $\mathfrak{s}_i = \mathrm{span}\{V_i, W_i\}$. The root space decomposition is:

$$\mathfrak{g} = \tau \oplus \{\mathfrak{l}_{ij} \mid i < j\} \oplus \{\mathfrak{k}_{ij} \mid i < j\} \oplus \{\mathfrak{s}_i \mid 1 \leq i \leq n\}.$$

Since $[V_i, W_i] = H_i$, the added dual roots are the following matrices (and their negatives):

$$\hat{\gamma}_i = \left[ \frac{V_i}{|V_i|}, \frac{W_i}{|W_i|} \right] = \frac{1}{2} H_i.$$

In addition to those of $so(2n)$, we have the following new sums and differences of dual roots that equal dual roots. In each case, a sample bracket value is provided:

$$\begin{array}{lll}
\hat{\alpha}_{ij} - \hat{\gamma}_i = -\hat{\gamma}_j & [\mathfrak{l}_{ij}, \mathfrak{s}_i] \subset \mathfrak{s}_j & [E_{ij}, V_i] = -V_j \\
\hat{\alpha}_{ij} + \hat{\gamma}_j = \hat{\gamma}_i & [\mathfrak{l}_{ij}, \mathfrak{s}_j] \subset \mathfrak{s}_i & [E_{ij}, V_j] = V_i \\
\hat{\beta}_{ij} - \hat{\gamma}_i = \hat{\gamma}_j & [\mathfrak{k}_{ij}, \mathfrak{s}_i] \subset \mathfrak{s}_j & [X_{ij}, V_i] = -W_j \\
\hat{\beta}_{ij} - \hat{\gamma}_j = \hat{\gamma}_i & [\mathfrak{k}_{ij}, \mathfrak{s}_j] \subset \mathfrak{s}_i & [X_{ij}, V_j] = W_i \\
\left.\begin{array}{l} \hat{\gamma}_i - \hat{\gamma}_j = \hat{\alpha}_{ij} \\ \hat{\gamma}_i + \hat{\gamma}_j = \hat{\beta}_{ij} \end{array}\right\} & [\mathfrak{s}_i, \mathfrak{s}_j] \subset \mathfrak{l}_{ij} \oplus \mathfrak{k}_{ij} & [V_i, V_j] = -\dfrac{1}{2}E_{ij} + \dfrac{1}{2}Y_{ij}.
\end{array}$$

Notice that $(\hat{\gamma}_i, \hat{\gamma}_j)$ is our first example of a pair of dual roots whose sum and difference *both* equal dual roots. Using Table 2, the new bracket relations (in addition to those of Table 3) are summarized in Table 4.

**Table 4.** The additional bracket relations for $\mathfrak{g} = so(2n+1)$.

| $[\cdot,\cdot]$ | $\mathbf{V_i}$ | $\mathbf{W_i}$ |
|---|---|---|
| $\mathbf{E_{ij}}$ | $-V_j$ | $-W_j$ |
| $\mathbf{F_{ij}}$ | $W_j$ | $-V_j$ |

| $[\cdot,\cdot]$ | $\mathbf{V_j}$ | $\mathbf{W_j}$ |
|---|---|---|
| $\mathbf{E_{ij}}$ | $V_i$ | $W_i$ |
| $\mathbf{F_{ij}}$ | $W_i$ | $-V_i$ |

| $[\cdot,\cdot]$ | $\mathbf{V_i}$ | $\mathbf{W_i}$ |
|---|---|---|
| $\mathbf{X_{ij}}$ | $-W_j$ | $-V_j$ |
| $\mathbf{Y_{ij}}$ | $V_j$ | $-W_j$ |

| $[\cdot,\cdot]$ | $\mathbf{V_j}$ | $\mathbf{W_j}$ |
|---|---|---|
| $\mathbf{X_{ij}}$ | $W_i$ | $V_i$ |
| $\mathbf{Y_{ij}}$ | $-V_i$ | $W_i$ |

| $[\cdot,\cdot]$ | $\mathbf{V_j}$ | $\mathbf{W_j}$ |
|---|---|---|
| $\mathbf{V_i}$ | $-\frac{1}{2}E_{ij} + \frac{1}{2}Y_{ij}$ | $\frac{1}{2}F_{ij} - \frac{1}{2}X_{ij}$ |
| $\mathbf{W_i}$ | $-\frac{1}{2}F_{ij} - \frac{1}{2}X_{ij}$ | $-\frac{1}{2}E_{ij} - \frac{1}{2}Y_{ij}$ |

## 8. The structure of $sp(n)$

Let $n > 0$ and $G = Sp(n)$, so $\mathfrak{g} = sp(n)$. Recall that the Lie algebra of the standard maximal torus of $G$ is:

$$\tau = \{\text{diag}(\theta_1 \mathbf{i}, ..., \theta_n \mathbf{i}) \mid \theta_i \in \mathbb{R}\}.$$

For each index $i$, let $H_i$ denote the diagonal matrix with $\mathbf{i}$ in position $(i,i)$ (and all other entries zero). Let $J_i$ denote the diagonal matrix with $\mathbf{j}$ in position $(i,i)$ (and all other entries zero), and let $K_i$ denote the diagonal matrix with $\mathbf{k}$ in position $(i,i)$. Notice that $H_i \in \tau$ and $J_i, K_i \in \tau^\perp$.

For each pair $(i,j)$ of distinct indices between 1 and $n$, let $E_{ij}$ denote the matrix with $+1$ in position $(i,j)$ and $-1$ in position $(j,i)$.

Let $F_{ij} \in \mathfrak{g}$ denote the matrix with $\mathbf{i}$ in positions $(i,j)$ and $(j,i)$. Let $A_{ij} \in \mathfrak{g}$ denote the matrix with $\mathbf{j}$ in positions $(i,j)$ and $(j,i)$. Let $B_{ij} \in \mathfrak{g}$ denote the matrix with $\mathbf{k}$ in positions $(i,j)$ and $(j,i)$.

The root space decomposition is:

$$\begin{aligned} sp(n) \quad &= \quad \tau \oplus \{\text{span}\{E_{ij}, F_{ij}\} \mid i<j\} \oplus \{\text{span}\{A_{ij}, B_{ij}\} \mid i<j\} \\ & \qquad \oplus\{\text{span}\{J_i, K_i\} \mid 1 \leq i \leq n\}. \end{aligned}$$

The dual roots are the following matrices (and their negatives):

$$\hat{\alpha}_{ij} = \left[\frac{E_{ij}}{|E_{ij}|}, \frac{F_{ij}}{|F_{ij}|}\right] = H_i - H_j$$

$$\hat{\beta}_{ij} = \left[\frac{A_{ij}}{|A_{ij}|}, \frac{B_{ij}}{|B_{ij}|}\right] = H_i + H_j$$

$$\hat{\gamma}_i = \left[\frac{J_i}{|J_i|}, \frac{K_i}{|K_i|}\right] = 2H_i.$$

Think of these dual roots initially as unrelated to the dual roots of $SO(2n+1)$ which bore the same names, but look for similarities. The only sums or differences of dual roots that equal dual roots are listed below, with sample bracket values provided:

$$\begin{array}{ll} \hat{\alpha}_{ij} + \hat{\alpha}_{jk} = \hat{\alpha}_{ik} & [E_{ij}, E_{jk}] = 2E_{ik} \\ \hat{\beta}_{ij} - \hat{\beta}_{jk} = \hat{\alpha}_{ik} & [A_{ij}, A_{jk}] = -2E_{ik} \\ \hat{\alpha}_{ij} + \hat{\beta}_{jk} = \hat{\beta}_{ik} & [E_{ij}, A_{jk}] = 2A_{ik} \\ \hat{\beta}_{ij} - \hat{\gamma}_i = -\hat{\alpha}_{ij} & [A_{ij}, J_i] = 2E_{ij} \\ \hat{\beta}_{ij} - \hat{\gamma}_j = \hat{\alpha}_{ij} & [A_{ij}, J_j] = -2E_{ij} \\ \hat{\alpha}_{ij} - \hat{\gamma}_i = -\hat{\beta}_{ij} & [A_{ij}, J_i] = -2A_{ij} \\ \hat{\alpha}_{ij} + \hat{\gamma}_j = \hat{\beta}_{ij} & [E_{ij}, J_j] = 2A_{ij}. \end{array}$$

We leave it to the reader in Exercise 11.4 to list all non-zero brackets of pairs of basis vectors, using the above sample values together with Table 2.

## 9. The Weyl group

Let $G$ be a compact path-connected Lie group with Lie algebra $\mathfrak{g}$. Let $T \subset G$ be a maximal torus with Lie algebra $\tau \subset \mathfrak{g}$. In this section, we will define and study the Weyl group of $G$, which can be thought of as a group of symmetries of the roots of $G$.

First, let $N(T)$ denote the **normalizer** of $T$, which means:

$$N(T) = \{g \in G \mid gTg^{-1} = T\}.$$

It is routine to check that $N(T)$ is a subgroup of $G$ and that $T$ is a normal subgroup of $N(T)$.

For each $g \in N(T)$, conjugation by $g$ is an automorphism of $T$, denoted $C_g : T \to T$. The derivative of $C_g$ at $I$ is the Lie algebra automorphism $\mathrm{Ad}_g : \tau \to \tau$. In fact, it is straightforward to see:

$$N(T) = \{g \in G \mid \mathrm{Ad}_g(\tau) = \tau\}.$$

One should expect automorphisms to preserve all of the fundamental structures of a Lie algebra, including its roots and dual roots.

**Proposition 11.15.** *For each $g \in N(T)$, $Ad_g : \tau \to \tau$ sends dual roots to dual roots.*

**Proof.** If $\hat{\alpha} \in \tau$ is a dual root with root space $\mathfrak{l} = \mathrm{span}\{E, F\}$, then $\mathrm{Ad}_g\hat{\alpha} \in \tau$ is a dual root with root space $\mathrm{Ad}_g(\mathfrak{l}) = \mathrm{span}\{\mathrm{Ad}_g E, \mathrm{Ad}_g F\}$. This is because for all $X \in \tau$,

$$\begin{aligned}[X, \mathrm{Ad}_g E] &= \mathrm{Ad}_g\left([\mathrm{Ad}_{g^{-1}}X, E]\right) = \mathrm{Ad}_g\left(\langle Ad_{g^{-1}}X, \hat{\alpha}\rangle \cdot F\right) \\ &= \langle Ad_{g^{-1}}X, \hat{\alpha}\rangle \cdot \mathrm{Ad}_g F = \langle X, \mathrm{Ad}_g\hat{\alpha}\rangle \cdot \mathrm{Ad}_g F.\end{aligned}$$

Similarly, $[X, \mathrm{Ad}_g F] = -\langle X, \mathrm{Ad}_g\hat{\alpha}\rangle \cdot \mathrm{Ad}_g E$, so $\mathrm{Ad}_g\hat{\alpha}$ is a dual root according to Definition 11.5. □

Since the conjugates of $T$ cover $G$, $N(T)$ is not all of $G$. In fact, we expect $N(T)$ to be quite small. The following shows at least that $N(T)$ is larger than $T$.

**Proposition 11.16.** *For each dual root, $\hat{\alpha}$, of $G$, there exists an element $g \in N(T)$ such that $Ad_g(\hat{\alpha}) = -\hat{\alpha}$, and $Ad_g(X) = X$ for all $X \in \tau$ with $X \perp \hat{\alpha}$.*

In other words, we can visualize $\mathrm{Ad}_g : \tau \to \tau$ as a reflection through the "hyperplane" $\hat{\alpha}^{\perp} = \{X \in \tau \mid X \perp \hat{\alpha}\}$.

**Proof.** Let $\hat{\alpha}$ be a dual root with root space $\mathfrak{l} = \operatorname{span}\{E, F\}$. Since $t \mapsto e^{tF}$ is a one-parameter group in $G$, $t \mapsto \mathrm{Ad}_{e^{tF}}$ is a one-parameter group of orthogonal automorphisms of $\mathfrak{g}$, with initial derivative equal to $\mathrm{ad}_F$. Notice that:

$$\text{For all } X \in \hat{\alpha}^{\perp},\ \mathrm{ad}_F(X) = -[X, F] = \langle X, \hat{\alpha}\rangle \cdot E = 0,$$

$$\mathrm{ad}_F(E) = -[E, F] = -\hat{\alpha} = -|\hat{\alpha}| \cdot \frac{\hat{\alpha}}{|\hat{\alpha}|},$$

$$\mathrm{ad}_F\left(\frac{\hat{\alpha}}{|\hat{\alpha}|}\right) = -\left[\frac{\hat{\alpha}}{|\hat{\alpha}|}, F\right] = \left\langle \frac{\hat{\alpha}}{|\hat{\alpha}|}, \hat{\alpha}\right\rangle \cdot E = |\hat{\alpha}| \cdot E.$$

Therefore $t \mapsto \mathrm{Ad}_{e^{tF}} = e^{\mathrm{ad}_{tF}}$ is a one-parameter group of orthogonal automorphisms of $\mathfrak{g}$ which acts as the identity on $\hat{\alpha}^{\perp} \subset \tau$ and which rotates $\operatorname{span}\left\{\frac{\hat{\alpha}}{|\hat{\alpha}|}, E\right\}$ with period $\frac{2\pi}{|\hat{\alpha}|}$. Thus, at time $t_0 = \frac{\pi}{|\hat{\alpha}|}$, the rotation is half complete, so it sends $\hat{\alpha} \mapsto -\hat{\alpha}$. Thus, the element $g = e^{t_0 F}$ lies in $N(T)$ and acts on $\tau$ as claimed in the proposition. ☐

We would like to think of $N(T)$ as a group of orthogonal automorphisms of $\tau$, but the problem is that different elements of $N(T)$ may determine the same automorphism of $\tau$:

**Lemma 11.17.** *For a pair $a, b \in N(T)$, $Ad_a = Ad_b$ on $\tau$ if and only if $a$ and $b$ lie in the same coset of $N(T)/T$.*

**Proof.**

$$\begin{aligned} \mathrm{Ad}_a = \mathrm{Ad}_b \text{ on } \tau \quad &\iff \quad C_a = C_b \text{ on } T \\ &\iff \quad C_{ab^{-1}} = \mathrm{I} \text{ on } T \\ &\iff \quad ab^{-1} \text{ commutes with every element of } T \\ &\iff \quad ab^{-1} \in T. \end{aligned}$$

☐

**Definition 11.18.** *The **Weyl group** of $G$ is $W(G) = N(T)/T$.*

So it is not $N(T)$ but $W(G)$ that should be thought of as a group of orthogonal automorphisms of $\tau$. Each $w = g{\cdot}T \in W(G)$ determines the automorphism of $\tau$ that sends $X \in \tau$ to

$$w \star X = \mathrm{Ad}_g X.$$

By the previous lemma, $w \star X$ is well-defined (independent of the coset representative $g \in N(T)$), and different elements of $W(G)$ determine different automorphisms of $\tau$.

**Proposition 11.19.** *$W(G)$ is finite.*

**Proof.** By the above remarks, $W(G)$ is isomorphic to a subgroup of the group of automorphisms of $\tau$. By Proposition 11.15, each $w \in W(G)$ determines a permutation of the $2m$ dual roots of $G$. If two elements $w_1, w_2 \in W$ determine the same permutation of the dual roots, then they determine the same linear map on the span of the dual roots. The proof of Proposition 11.7 shows that the span of the dual roots equals the orthogonal complement in $\tau$ of the Lie algebra of the center of $G$. Since each element of $W(G)$ acts as the identity on the Lie algebra of the center of $G$, this shows that $w_1$ and $w_2$ determine the same automorphism of $\tau$, and therefore $w_1 = w_2$. Thus, different elements of $W(G)$ must determine different permutations of the dual roots. It follows that $W(G)$ is isomorphic to a subgroup of the group of permutations of the $2m$ dual roots and thus has finite order that divides $(2m)!$ ☐

Proposition 11.16 guarantees that for each dual root $\hat{\alpha}$, there exists an element $w_{\hat{\alpha}} \in W(G)$ such that $w_{\hat{\alpha}} \star \hat{\alpha} = -\hat{\alpha}$ and $w_{\hat{\alpha}} \star X = X$ for all $X \in \hat{\alpha}^{\perp}$. It turns out that such elements generate $W(G)$:

**Proposition 11.20.** *Every element of $W(G)$ equals a product of finitely many of the $w_{\hat{\alpha}}$'s.*

We will not prove this proposition. It implies that $W(G)$ depends only on the Lie algebra. That is, if two Lie groups have isomorphic Lie algebras, then they have isomorphic Weyl groups. By contrast, the normalizer of the maximal torus of $SO(3)$ is not isomorphic to that of $Sp(1)$, even though $so(3) \cong sp(1)$ (see Exercises 9.10 and 9.11 for descriptions of these normalizers).

It is useful to derive an explicit formula for the reflection through the hyperplane $\hat{\alpha}^{\perp}$:

**Lemma 11.21.** *If $\hat{\alpha}$ is a dual root, then $w_{\hat{\alpha}} \star X = X - 2\frac{\langle \hat{\alpha}, X\rangle}{\langle \hat{\alpha}, \hat{\alpha}\rangle}\hat{\alpha}$ for all $X \in \tau$.*

**Proof.** $X$ uniquely decomposes as the sum of a vector parallel to $\hat{\alpha}$ and a vector perpendicular to $\hat{\alpha}$ in the following explicit manner:

$$X = X_{\|} + X_{\perp} = \left(\frac{\langle\hat{\alpha}, X\rangle}{\langle\hat{\alpha},\hat{\alpha}\rangle}\hat{\alpha}\right) + \left(X - \frac{\langle\hat{\alpha}, X\rangle}{\langle\hat{\alpha},\hat{\alpha}\rangle}\hat{\alpha}\right).$$

We have $w_{\hat{\alpha}}(X) = -X_{\|} + X_{\perp} = X - 2\frac{\langle\hat{\alpha},X\rangle}{\langle\hat{\alpha},\hat{\alpha}\rangle}\hat{\alpha}$. □

**Proposition 11.22.** *$W(SU(n))$ is isomorphic to $S_n$, the group of all permutations of $n$ objects.*

**Proof.** Using Lemma 11.21, one can check that $w_{\hat{\alpha}_{ij}} \star X$ is obtained from $X \in \tau$ by exchanging the $i^{\text{th}}$ and $j^{\text{th}}$ diagonal entries. For example,

$$w_{\hat{\alpha}_{12}} \star \operatorname{diag}(\lambda_1\mathbf{i}, \lambda_2\mathbf{i}, \lambda_3\mathbf{i}, ..., \lambda_n\mathbf{i}) = \operatorname{diag}(\lambda_2\mathbf{i}, \lambda_1\mathbf{i}, \lambda_3\mathbf{i}, ..., \lambda_n\mathbf{i}).$$

The collection $\{w_{\hat{\alpha}_{ij}}\}$ generates the group, $S_n$, of all permutations of the $n$ diagonal entries, so Proposition 11.20 implies that $W(SU(n))$ is isomorphic to $S_n$. □

An explicit coset representative, $g_{ij} \in N(T) \subset SU(n)$, for each $w_{\hat{\alpha}_{ij}}$ can be found using the construction in the proof of Proposition 11.16. For example, in $G = SU(3)$, we can choose:

$$g_{23} = e^{(\pi/2)E_{23}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \text{ or } g_{23} = e^{(\pi/2)F_{23}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & \mathbf{i} \\ 0 & \mathbf{i} & 0 \end{pmatrix}.$$

Finally, we will determine the Weyl groups of the remaining classical groups. For each of $G \in \{SO(2n), SO(2n+1), Sp(n)\}$, we previously chose a basis of $\tau$, which in all three cases was denoted $\{H_1, ..., H_n\}$. These basis elements are mutually orthogonal and have the same length, $l$. They are tangent to the circles that comprise $T$, so they generate one-parameter groups, $t \mapsto e^{tH_i}$, with period $2\pi$. In fact, $\{\pm H_1, ..., \pm H_n\}$ are the only vectors in $\tau$ of length $l$ that generate one-parameter groups with period $2\pi$. For any $g \in N(T)$, $\operatorname{Ad}_g : \tau \to \tau$ must preserve this property and therefore must permute the set $\{\pm H_1, ..., \pm H_n\}$. We will think of $W(G)$ as a group of permutations of this set (rather than of the set of dual roots).

**Proposition 11.23.** $|W(SO(2n+1))| = |W(Sp(n))| = 2^n n!$, *and* $|W(SO(2n))| = 2^{n-1}n!$

**Proof.** For each of $G \in \{SO(2n), SO(2n+1), Sp(n)\}$, there are dual roots denoted $\hat{\alpha}_{ij}$ and $\hat{\beta}_{ij}$. Using Lemma 11.21, the corresponding Weyl group elements permute the set $\{\pm H_1, ..., \pm H_n\}$ as follows:

$$w_{\hat{\alpha}_{ij}} \text{ sends } H_i \mapsto H_j, \quad H_j \mapsto H_i, \quad H_k \mapsto H_k \text{ for all } k \notin \{i, j\},$$
$$w_{\hat{\beta}_{ij}} \text{ sends } H_i \mapsto -H_j, \quad H_j \mapsto -H_i, \quad H_k \mapsto H_k \text{ for all } k \notin \{i, j\}.$$

For $G \in \{SO(2n+1), Sp(n)\}$ we additionally have dual roots denoted $\{\hat{\gamma}_i\}$ which give the following permutations:

$$w_{\hat{\gamma}_i} \text{ sends } H_i \mapsto -H_i, \quad H_k \mapsto H_k \text{ for all } k \neq i.$$

By Proposition 11.20, the Weyl group is isomorphic to the group of permutations of set $\{\pm H_1, ..., \pm H_n\}$ generated by the above permutations. For $G \in \{SO(2n+1), Sp(n)\}$, one can generate any permutation of the $n$ indices together with any designation of which of the $n$ indices become negative, giving $2^n n!$ possibilities. For $G = SO(2n)$, the number of negative indices must be even (check that this is the only restriction), so there are half as many total possibilities. $\square$

## 10. Towards the classification theorem

In this section, we very roughly indicate the proof of the previously mentioned classification theorem for compact Lie groups which stated:

**Theorem 11.24.** *The Lie algebra of every compact Lie group, $G$, is isomorphic to the Lie algebra of a product $G_1 \times G_2 \times \cdots \times G_k$, where each $G_i$ is one of $\{SO(n), SU(n), Sp(n)\}$ for some $n$ or is one of the five exceptional Lie groups: $G_2, F_4, E_6, E_7$ and $E_8$.*

It suffices to prove this theorem assuming that $G$ has a finite center, so we'll henceforth assume that all of our compact Lie groups have finite centers. In this case, the dual roots of $G$ are a finite collection of vectors in $\tau$ which form a "root system" according to the following definition:

**Definition 11.25.** *Let $\tau$ be a real vector space that has an inner product, $\langle \cdot, \cdot \rangle$. Let $R$ be a finite collection of non-zero vectors in $\tau$ which span $\tau$. The pair $(\tau, R)$ is called a* ***root system*** *if the following properties are satisfied:*

(1) *If $\alpha \in R$, then $-\alpha \in R$, but no other multiple of $\alpha$ is in $R$.*

(2) *If* $\alpha, \beta \in R$, *then* $w_\alpha \star \beta = \beta - 2\frac{\langle\beta,\alpha\rangle}{\langle\alpha,\alpha\rangle}\alpha \in R$.

(3) *If* $\alpha, \beta \in R$, *then the quantity* $2\frac{\langle\beta,\alpha\rangle}{\langle\alpha,\alpha\rangle}$ *is an integer.*

*In this case, the elements of $R$ are called **roots**, and the dimension of $\tau$ is called the **rank** of the root system.*

In property (2), each $a \in \tau$ determines the orthogonal endomorphism of $\tau$ that sends $X \in \tau$ to the vector $w_a \star X \in \tau$ defined as $w_a \star X = X - 2\frac{\langle X,a\rangle}{\langle a,a\rangle}a$. That is, $w_a : \tau \to \tau$ is the reflection through the hyperplane $a^\perp$. Property (2) says that for each root $\alpha$, the reflection $w_\alpha$ sends roots to roots. The **Weyl group** of $(\tau, R)$, denoted $W(\tau, R)$, is defined as the group of all endomorphisms of $\tau$ obtained by composing a finite number of the $w_\alpha$'s. As before, $W(\tau, R)$ is isomorphic to a subgroup of the group of permutations of $R$.

Some representation theory is required to prove that the dual roots of $G$ satisfy property (3). Interpreting $\frac{\langle\beta,\alpha\rangle}{\langle\alpha,\alpha\rangle}$ as in the proof of Lemma 11.21, property (3) says that the projection of $\beta$ onto $\alpha$ (previously denoted $\beta_{||}$) must be an integer or half-integer multiple of $\alpha$, and vice-versa. This implies very strong restrictions on the angle $\angle(\alpha, \beta)$ and on the ratio $\frac{|\alpha|}{|\beta|}$. In particular, the following is straightforward to prove using only property (3):

**Proposition 11.26.** *Let $(\tau, R)$ be a root system. If $\alpha, \beta \in R$, then one of the following holds:*

(0) $\langle\alpha, \beta\rangle = 0$.

(1) $|\alpha| = |\beta|$ *and* $\angle(\alpha, \beta) \in \{60°, 120°\}$.

(2) $\max\{|\alpha|, |\beta|\} = \sqrt{2}\cdot\min\{|\alpha|, |\beta|\}$ *and* $\angle(\alpha, \beta) \in \{45°, 135°\}$.

(3) $\max\{|\alpha|, |\beta|\} = \sqrt{3}\cdot\min\{|\alpha|, |\beta|\}$ *and* $\angle(\alpha, \beta) \in \{30°, 150°\}$.

In fact, the definition of a root system is so restrictive, root systems have been completely classified:

**Theorem 11.27.** *Every root system is equivalent to the system of dual roots for a Lie group of the form $G = G_1 \times G_2 \times \cdots \times G_k$, where each $G_i$ is one of $\{SO(n), SU(n), Sp(n)\}$ for some $n$ or is one of the five exceptional Lie groups.*

In order for this classification of root systems to yield a proof of Theorem 11.24, it remains only to establish that:

**Theorem 11.28.** *Two compact Lie groups with equivalent systems of dual roots must have isomorphic Lie algebras.*

The notion of equivalence in the previous two theorems is formalized as follows:

**Definition 11.29.** *The root system $(\tau, R)$ is said to be **equivalent** to the root system $(\tau', R')$ if there exists a linear isomorphism $f : \tau \to \tau'$ that sends $R$ onto $R'$ such that for all $\alpha \in R$ and $X \in \tau$ we have:*

$$f(w_\alpha \star X) = w_{f(\alpha)} \star f(X).$$

If $f$ is orthogonal (meaning that $\langle f(X), f(Y)\rangle = \langle X, Y\rangle$ for all $X, Y \in \tau$), then the hyperplane-reflection property in this definition is automatic. An example of a non-orthogonal equivalence is given in Exercise 11.22.

The proofs of Theorems 11.27 and 11.28 are difficult; see [9] for complete details. One of the key steps in the proof of Theorem 11.27 involves showing that every root system contains a special type of basis called a "base," defined as follows:

**Definition 11.30.** *Let $(\tau, R)$ be a root system, and let $\Delta \subset R$ be a collection of the roots that form a basis of $\tau$, which implies that every $\alpha \in R$ can be written uniquely as a linear combination of elements of $\Delta$. We call $\Delta$ a **base** of $R$ if the non-zero coefficients in each such linear combination are integers and are either all positive (in which case $\alpha$ is called a **positive root**) or all negative (in which case $\alpha$ is called a **negative root**).*

For a proof that every root system has a base, see [**7**] or [**9**]. The most natural base for the system of dual roots of $G = SU(n)$ is:

$$\Delta = \{\hat\alpha_{12}, \hat\alpha_{23}, ..., \hat\alpha_{(n-1)n}\}.$$

This base induces the same notion of "positive" that was provided in Section 1; namely, $\hat\alpha_{ij}$ is positive if and only if $i < j$. For example, $\hat\alpha_{25}$ is positive because $\hat\alpha_{25} = \hat\alpha_{23} + \hat\alpha_{34} + \hat\alpha_{45}$.

A natural base for the system of dual roots of $G = SO(2n)$ is:

$$\Delta = \{\hat{\alpha}_{12}, \hat{\alpha}_{23}, ..., \hat{\alpha}_{(n-1)n}, \hat{\beta}_{(n-1)n}\}.$$

As before, $\hat{\alpha}_{ij}$ is positive if and only if $i < j$. Also, $\hat{\beta}_{ij}$ is positive and $-\hat{\beta}_{ij}$ is negative for each pair $(i, j)$. For example, when $n = 8$, so $G = SO(16)$, we can verify that $\hat{\beta}_{35}$ is positive by writing:

$$\hat{\beta}_{35} = \hat{\alpha}_{34} + \hat{\alpha}_{45} + 2\hat{\alpha}_{56} + 2\hat{\alpha}_{67} + \hat{\alpha}_{78} + \hat{\beta}_{78}.$$

A base for the system of dual roots of $G = SO(2n + 1)$ is:

$$\Delta = \{\hat{\alpha}_{12}, \hat{\alpha}_{23}, ..., \hat{\alpha}_{(n-1)n}, \hat{\gamma}_n\}.$$

As before, $\hat{\alpha}_{ij}$ is positive if and only if $i < j$, each $\hat{\beta}_{ij}$ is positive, and each $-\hat{\beta}_{ij}$ is negative. Further, $\hat{\gamma}_i$ is positive and $-\hat{\gamma}_i$ is negative for each index $1 \leq i \leq n$. For example, when $n = 8$, so $G = SO(17)$, we can verify that $\hat{\beta}_{35}$ and $\hat{\gamma}_3$ are positive by writing:

$$\hat{\beta}_{35} = \hat{\alpha}_{34} + \hat{\alpha}_{45} + 2\hat{\alpha}_{56} + 2\hat{\alpha}_{67} + 2\hat{\alpha}_{78} + 2\hat{\gamma}_8,$$
$$\hat{\gamma}_3 = \hat{\alpha}_{34} + \hat{\alpha}_{45} + \hat{\alpha}_{56} + \hat{\alpha}_{67} + \hat{\alpha}_{78} + \hat{\gamma}_8.$$

A base for the system of dual roots of $G = Sp(n)$ is:

$$\Delta = \{\hat{\alpha}_{12}, \hat{\alpha}_{23}, ..., \hat{\alpha}_{(n-1)n}, \hat{\gamma}_n\}.$$

As before, $\hat{\alpha}_{ij}$ is positive if and only if $i < j$, each $\hat{\beta}_{ij}$ and each $\hat{\gamma}_i$ is positive, and each $-\hat{\beta}_{ij}$ and each $-\hat{\gamma}_i$ is negative. For example, in $G = Sp(8)$, we can verify that $\hat{\beta}_{35}$ and $\hat{\gamma}_3$ are positive by writing:

$$\hat{\beta}_{35} = \hat{\alpha}_{34} + \hat{\alpha}_{45} + 2\hat{\alpha}_{56} + 2\hat{\alpha}_{67} + 2\hat{\alpha}_{78} + \hat{\gamma}_8,$$
$$\hat{\gamma}_3 = 2\hat{\alpha}_{34} + 2\hat{\alpha}_{45} + 2\hat{\alpha}_{56} + 2\hat{\alpha}_{67} + 2\hat{\alpha}_{78} + \hat{\gamma}_8.$$

For a compact Lie group, $G$, each root space, $\mathfrak{l}$, is associated with two dual roots. A base, $\Delta$, will designate one of them as positive (denoted $\hat{\alpha}$) and the other as negative (denoted $-\hat{\alpha}$). Therefore, a base provides a notion of "clockwise" for each $\mathfrak{l}$; namely, clockwise with respect to an ordered orthonormal basis $\{E, F\}$ of $\mathfrak{l}$ such that $[E, F]$ equals the positive dual root. None of the above bases for the classical groups are unique, which reflects the lack of a canonical notion of clockwise for the individual root spaces. A different base would induce a different division of the roots into positive and negative roots, and thus different notions of clockwise for the root spaces.

**Lemma 11.31.** *If $(\tau, R)$ is a root system, $\Delta$ is a base, and $\alpha, \beta \in \Delta$, then one of the following holds:*

(0) $\langle \alpha, \beta \rangle = 0$.

(1) $|\alpha| = |\beta|$ *and* $\angle(\alpha, \beta) = 120°$.

(2) $max\{|\alpha|, |\beta|\} = \sqrt{2} \cdot min\{|\alpha|, |\beta|\}$, *and* $\angle(\alpha, \beta) = 135°$.

(3) $max\{|\alpha|, |\beta|\} = \sqrt{3} \cdot min\{|\alpha|, |\beta|\}$, *and* $\angle(\alpha, \beta) = 150°$.

**Proof.** By Proposition 11.26, we need only prove that $\angle(\alpha, \beta)$ is not acute. For each of the three possible acute angles, it is straightforward to show that $w_\alpha \star \beta = \alpha - \beta$ or $w_\beta \star \alpha = \beta - \alpha$. In either case, $\alpha - \beta \in R$ is a root whose unique expression as a linear combination of elements from $\Delta$ has a positive and a negative coefficient, contradicting the definition of base. □

It turns out that to determine the equivalence class of a root system $(\tau, R)$, one only needs to know the angles between pairs of vectors from a base, $\Delta$, of the root system. A **Dynkin diagram** is a graph that encodes exactly this information. The nodes of the Dynkin diagram are the elements of $\Delta$ (so the number of nodes equals the rank of the root system). For a pair of nodes representing elements $\alpha, \beta \in \Delta$, we put 0, 1, 2, or 3 edges between them to represent the possibilities enumerated in Lemma 11.31. Further, we decorate each double or triple edge with an arrow from the vertex associated with the longer root towards the vertex associated with the smaller root. It can be proven that the Dynkin diagram does not depend on the choice of base and that it determines the equivalence class of the root system. The classification of root systems was achieved by classifying all possible Dynkin diagrams. The Dynkin diagrams for systems of dual roots of the classical groups are pictured in Figure 1.

## 11. Complexified Lie algebras

In this section, we will define the "complexification" of a Lie algebra, to build a bridge between this book and more advanced books which typically emphasize roots of a complexified Lie algebra.

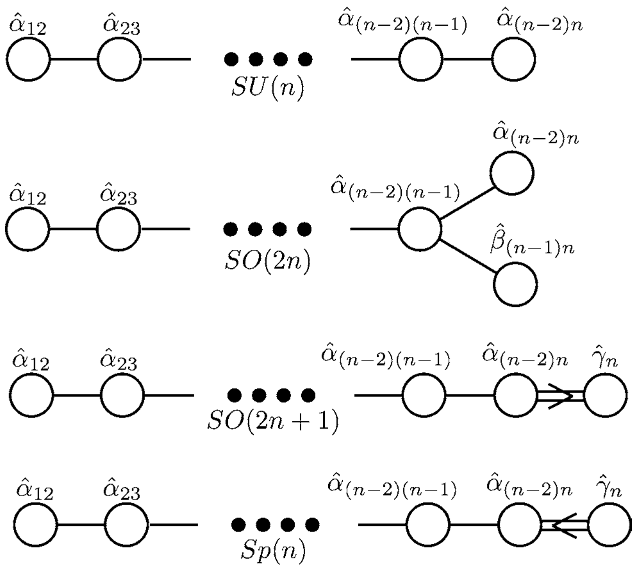



**Figure 1.** The Dynkin diagrams of the classical Lie groups.

**Definition 11.32.** *Let $V$ be an $n$-dimensional vector space over $\mathbb{R}$. The* ***complexification*** *of $V$ is defined as:*

$$V_{\mathbb{C}} = \{X + Y\mathbf{i} \mid X, Y \in V\}.$$

Notice that $V_{\mathbb{C}}$ is an $n$-dimensional vector space over $\mathbb{C}$, with vector addition and scalar multiplication defined in the obvious way:

$$(X_1 + Y_1\mathbf{i}) + (X_2 + Y_2\mathbf{i}) = (X_1 + X_2) + (Y_1 + Y_2)\mathbf{i},$$
$$(a + b\mathbf{i}) \cdot (X + Y\mathbf{i}) = (a \cdot X - b \cdot Y) + (b \cdot X + a \cdot Y)\mathbf{i},$$

for all $X, X_1, X_2, Y, Y_1, Y_2 \in V$ and $a, b \in \mathbb{R}$.

If $V$ has an inner product, $\langle \cdot, \cdot \rangle$, then this induces a natural complex-valued inner product on $V_{\mathbb{C}}$ defined as:

$$\langle X_1 + Y_1\mathbf{i}, X_2 + Y_2\mathbf{i}\rangle_{\mathbb{C}} = (\langle X_1, X_2\rangle + \langle Y_1, Y_2\rangle) + (\langle Y_1, X_2\rangle - \langle X_1, Y_2\rangle)\mathbf{i},$$

which is designed to satisfy all of the familiar properties of the standard hermitian inner product on $\mathbb{C}^n$ enumerated in Proposition 3.3.

If $G$ is a Lie group with Lie algebra $\mathfrak{g}$, then $\mathfrak{g}_{\mathbb{C}}$ inherits a "complex Lie bracket" operation defined in the most natural way:

$$[X_1 + Y_1\mathbf{i}, X_2 + Y_2\mathbf{i}]_{\mathbb{C}} = ([X_1, X_2] - [Y_1, Y_2]) + ([X_1, Y_2] + [Y_1, X_2])\mathbf{i}.$$

This operation satisfies the familiar Lie bracket properties from Proposition 8.4 (with scalars $\lambda_1, \lambda_2 \in \mathbb{C}$), including the Jacobi identity.

Potential confusion arises when $G \subset GL(n, \mathbb{C})$ or $GL(n, \mathbb{H})$, since the symbol "$\mathbf{i}$" already has a meaning for the entries of matrices in $\mathfrak{g}$. In these cases, one should choose a different (initially unrelated) symbol, like "$\mathbf{I}$", for denoting elements of $\mathfrak{g}_{\mathbb{C}}$. See Exercises 11.20 and 11.21 for descriptions of $so(n)_{\mathbb{C}}$ and $u(n)_{\mathbb{C}}$.

If $G$ is a compact Lie group, then the root space decomposition, $\mathfrak{g} = \tau \oplus \mathfrak{l}_1 \oplus \cdots \oplus \mathfrak{l}_m$, induces a decomposition of $\mathfrak{g}_{\mathbb{C}}$ which is orthogonal with respect to $\langle \cdot, \cdot \rangle_{\mathbb{C}}$:

$$\mathfrak{g}_{\mathbb{C}} = \tau_{\mathbb{C}} \oplus (\mathfrak{l}_1)_{\mathbb{C}} \oplus \cdots \oplus (\mathfrak{l}_m)_{\mathbb{C}}.$$

Notice that $\tau_{\mathbb{C}}$ is an abelian $\mathbb{C}$-subspace of $\mathfrak{g}_{\mathbb{C}}$ ("abelian" means that every pair of vectors in $\tau_{\mathbb{C}}$ brackets to zero) and is maximal in the sense that it is not contained in any larger abelian $\mathbb{C}$-subspace of $\mathfrak{g}_{\mathbb{C}}$.

Each root space $\mathfrak{l}_i$ is associated with two roots, called $\alpha_i$ and $-\alpha_i$. We intend to further decompose each $(\mathfrak{l}_i)_{\mathbb{C}}$ into two 1-dimensional $\mathbb{C}$-subspaces, one for each of these two roots. That is, we will write:

$$(\mathfrak{l}_i)_{\mathbb{C}} = \mathfrak{g}_{\overline{\alpha}_i} \oplus \mathfrak{g}_{\overline{-\alpha_i}}, \tag{11.8}$$

with this notation defined as follows:

**Definition 11.33.** *If $\alpha$ is a root of $G$ and $\{E, F\}$ is an orthonormal basis of the corresponding root space, $\mathfrak{l}$, ordered so that the corresponding dual root is $\hat{\alpha} = [E, F]$, then:*

(1) *Define the $\mathbb{C}$-linear function $\overline{\alpha} : \tau_{\mathbb{C}} \to \mathbb{C}$ so that for all $X = X_1 + X_2\mathbf{i} \in \tau_{\mathbb{C}}$, we have*

$$\overline{\alpha}(X) = (-\mathbf{i}) \cdot (\alpha(X_1) + \alpha(X_2)\mathbf{i}) = \alpha(X_2) - \alpha(X_1)\mathbf{i}.$$

(2) *Define $\mathfrak{g}_{\overline{\alpha}} = span_{\mathbb{C}}\{E + F\mathbf{i}\} = \{\lambda(E + F\mathbf{i}) \mid \lambda \in \mathbb{C}\} \subset \mathfrak{l}_{\mathbb{C}}$.*

If $\{E, F\}$ is a correctly ordered basis for $\alpha$, then one for $-\alpha$ is $\{F, E\}$ or $\{E, -F\}$. In Equation 11.8, notice that $\mathfrak{g}_{\overline{\alpha}} = \text{span}_{\mathbb{C}}\{E + F\mathbf{i}\}$ and $\mathfrak{g}_{\overline{-\alpha}} = \text{span}_{\mathbb{C}}\{F + E\mathbf{i}\}$ are orthogonal with respect to $\langle \cdot, \cdot \rangle_{\mathbb{C}}$.

The space $\mathfrak{g}_{\overline{\alpha}}$ is well-defined, meaning independent of the choice of basis $\{E, F\}$. To see this, notice that another correctly ordered basis would look like $\{R_\theta E, R_\theta F\}$, where $R_\theta$ denotes a counterclockwise rotation of $\mathfrak{l}$ through angle $\theta$ (this assertion is justified in Exercise 11.7). Setting $\lambda = e^{-i\theta} = \cos\theta - \mathbf{i}\sin\theta$ gives:

$$\begin{aligned} \lambda \cdot (E + F\mathbf{i}) &= ((\cos\theta)E + (\sin\theta)F) + ((\cos\theta)F - (\sin\theta)E)\mathbf{i} \\ &= (R_\theta E) + (R_\theta F)\mathbf{i}. \end{aligned}$$

Thus, $\text{span}_{\mathbb{C}}\{E + F\mathbf{i}\} = \text{span}_{\mathbb{C}}\{(R_\theta E) + (R_\theta F)\mathbf{i}\}$.

The motivation for Definition 11.33 is the following:

**Proposition 11.34.** *If $\alpha$ is a root of $G$, then for each $X \in \tau_{\mathbb{C}}$, the value $\overline{\alpha}(X) \in \mathbb{C}$ is an eigenvalue of the function $ad_X : \mathfrak{g}_{\mathbb{C}} \to \mathfrak{g}_{\mathbb{C}}$ (which sends $V \mapsto [X, V]_{\mathbb{C}}$), and each vector in $\mathfrak{g}_{\overline{\alpha}}$ is a corresponding eigenvector.*

**Proof.** By $\mathbb{C}$-linearity, it suffices to verify this for $X \in \tau$, which is done as follows:

$$\begin{aligned} [X, E + F\mathbf{i}]_{\mathbb{C}} &= [X, E] + [X, F]\mathbf{i} = \alpha(X)(F - E\mathbf{i}) \\ &= (-\mathbf{i} \cdot \alpha(X))(E + F\mathbf{i}) = \overline{\alpha}(X) \cdot (E + F\mathbf{i}). \end{aligned}$$

□

For our general definition, we will use:

**Definition 11.35.** *A non-zero $\mathbb{C}$-linear function $\omega : \tau_{\mathbb{C}} \to \mathbb{C}$ is called a **complex root** of $\mathfrak{g}_{\mathbb{C}}$ if there exists a non-zero $\mathbb{C}$-subspace $\mathfrak{g}_\omega \subset \mathfrak{g}_{\mathbb{C}}$ (called a **complex root space**) such that for all $X \in \tau_{\mathbb{C}}$ and all $V \in \mathfrak{g}_\omega$ we have:*

$$[X, V]_{\mathbb{C}} = \omega(X) \cdot V.$$

*In other words, the elements of $\mathfrak{g}_\omega$ are eigenvectors of $ad_X$ for each $X \in \tau_{\mathbb{C}}$, and $\omega$ catalogs the corresponding eigenvalues.*

The notations "$\mathfrak{g}_\omega$" and "$\mathfrak{g}_{\overline{\alpha}}$" are consistent because of:

**Proposition 11.36.** *If $\alpha$ is a root of $G$, then $\overline{\alpha}$ is a complex root of $\mathfrak{g}_{\mathbb{C}}$, and all complex roots of $\mathfrak{g}_{\mathbb{C}}$ come from roots of $G$ in this way.*

*Thus, $\mathfrak{g}_{\mathbb{C}}$ decomposes uniquely as an orthogonal direct sum of complex root spaces:*

$$\begin{aligned}\mathfrak{g}_{\mathbb{C}} &= \tau_{\mathbb{C}} \oplus \{\mathfrak{g}_\omega \mid \omega \textit{ is a complex root of } \mathfrak{g}_{\mathbb{C}}\}\\ &= \tau_{\mathbb{C}} \oplus \{\mathfrak{g}_{\overline{\alpha}} \mid \alpha \textit{ is a root of } G\}\\ &= \tau_{\mathbb{C}} \oplus \mathfrak{g}_{\overline{\alpha}_1} \oplus \mathfrak{g}_{-\overline{\alpha}_1} \oplus \cdots \oplus \mathfrak{g}_{\overline{\alpha}_m} \oplus \mathfrak{g}_{-\overline{\alpha}_m}.\end{aligned}$$

One advantage of the complex setting is that $\overline{\alpha}_i$ and $-\overline{\alpha}_i$ correspond to different complex root spaces, so complex roots correspond one-to-one with complex root spaces. Another advantage is that the following complexified version of Theorem 11.13 has a short proof:

**Lemma 11.37.** *Suppose that $\omega_1$ and $\omega_2$ are complex roots of $\mathfrak{g}_{\mathbb{C}}$. If $V_1 \in \mathfrak{g}_{\omega_1}$ and $V_2 \in \mathfrak{g}_{\omega_2}$, then*

$$[V_1, V_2]_{\mathbb{C}} \in \begin{cases} \tau & \textit{if } \omega_1 = -\omega_2 \\ \mathfrak{g}_{\omega_1+\omega_2} & \textit{if } \omega_1 + \omega_2 \textit{ is a complex root of } \mathfrak{g}_{\mathbb{C}} \\ \{0\} & \textit{otherwise.} \end{cases}$$

**Proof.** Omitting the "$\mathbb{C}$" subscripts of Lie brackets for clarity, the complexified version of the Jacobi identity gives that for all $X \in \tau_{\mathbb{C}}$:

$$\begin{aligned}[X,[V_1,V_2]] &= -[V_1,[V_2,X]] - [V_2,[X,V_1]]\\ &= -[[X,V_2],V_1] + [[X,V_1],V_2]\\ &= (\omega_1(X) + \omega_2(X))[V_1,V_2],\end{aligned}$$

from which the three cases follow. □

**Alternative proof of Theorem 11.13.** For distinct indices $i, j$,

$$\mathfrak{g}_{\pm\overline{\alpha}_i} = \mathrm{span}_{\mathbb{C}}\{E_i \pm F_i\mathbf{i}\} \text{ and } \mathfrak{g}_{\pm\overline{\alpha}_j} = \mathrm{span}_{\mathbb{C}}\{E_j \pm F_j\mathbf{i}\}.$$

Lemma 11.37 says that the following two brackets:

$$[E_i + F_i\mathbf{i}, E_j + F_j\mathbf{i}]_{\mathbb{C}} = ([E_i,E_j] - [F_i,F_j]) + ([E_i,F_j] + [F_i,E_j])\mathbf{i},$$
$$[E_i + F_i\mathbf{i}, E_j - F_j\mathbf{i}]_{\mathbb{C}} = ([E_i,E_j] + [F_i,F_j]) + (-[E_i,F_j] + [F_i,E_j])\mathbf{i},$$

lie respectively in $\mathfrak{g}_{\overline{\alpha}_i+\overline{\alpha}_j}$ and $\mathfrak{g}_{\overline{\alpha}_i-\overline{\alpha}_j}$. The convention here is that $\mathfrak{g}_\omega = \{0\}$ if $\omega$ is not a complex root. Write $\mathfrak{l}^+_{ij} = \mathrm{span}\{E^+_{ij}, F^+_{ij}\}$ and

$\mathfrak{l}_{ij}^- = \text{span}\{E_{ij}^-, F_{ij}^-\}$, where these basis vectors may be zero. For the sum of the above two vectors, we have:

$$\begin{aligned} 2[E_i, E_j] + 2[F_i, E_j]\mathbf{i} &\in \mathfrak{g}_{\overline{\alpha}_i + \overline{\alpha}_j} \oplus \mathfrak{g}_{\overline{\alpha}_i - \overline{\alpha}_j} \\ &= \mathfrak{g}_{\overline{\alpha_i + \alpha_j}} \oplus \mathfrak{g}_{\overline{\alpha_i - \alpha_j}} \\ &= \text{span}_{\mathbb{C}}\{E_{ij}^+ + F_{ij}^+\mathbf{i}\} \oplus \text{span}_{\mathbb{C}}\{E_{ij}^- + F_{ij}^-\mathbf{i}\}. \end{aligned}$$

Thus,

$$2[E_i, E_j] \in \text{span}\{E_{ij}^+, F_{ij}^+, E_{ij}^-, F_{ij}^-\} = \mathfrak{l}_{ij}^+ \oplus \mathfrak{l}_{ij}^-.$$

□

## 12. Exercises

Unless specified otherwise, assume that $G$ is a compact path-connected Lie group with Lie algebra $\mathfrak{g}$, and $T \subset G$ is a maximal torus with Lie algebra $\tau \subset \mathfrak{g}$.

**Ex. 11.1.** For each $G \in \{SU(n), SO(2n), SO(2n+1), Sp(n)\}$, how many roots does $G$ have?

**Ex. 11.2.** For any linear function $\alpha : \tau \to \mathbb{R}$, prove there exists a unique vector $\hat{\alpha} \in \tau$ such that for all $X \in \tau$, $\alpha(X) = \langle \hat{\alpha}, X \rangle$.

*Hint: Define $\hat{\alpha}$ in terms of an orthonormal basis of $\tau$.*

**Ex. 11.3.** If one begins with a different maximal torus of $G$, show that this does not affect the equivalence class of the system of dual roots of $G$ or the isomorphism class of $W(G)$.

**Ex. 11.4.** Create tables describing all non-zero brackets of basis elements of $sp(n)$, as was done in this chapter for the other classical groups.

**Ex. 11.5.** If $G = G_1 \times G_2$, describe the roots and dual roots and Weyl group of $G$ in terms of those of $G_1$ and $G_2$.

**Ex. 11.6.** If $\hat{\alpha}$ is a dual root with root space $\mathfrak{l} = \text{span}\{E, F\}$, prove that $\text{span}\{E, F, \hat{\alpha}\}$ is a subalgebra of $\mathfrak{g}$ which is isomorphic to $su(2)$.

**Ex. 11.7.** If $\{E, F\}$ is an ordered orthonormal basis of the root space $\mathfrak{l}$, then any other ordered orthonormal basis of $\mathfrak{l}$ will be of the form $\{E' = L_g E, F' = L_g F\}$ for some $g \in O(2)$. Show that $\hat{\alpha} = [E, F]$ and $\hat{\alpha}' = [E', F']$ are equal if and only if $g \in SO(2)$; otherwise $\hat{\alpha}' = -\hat{\alpha}$.

**Ex. 11.8.** Prove that the center of $G$ equals the intersection of all maximal tori of $G$.

**Ex. 11.9.** Define the **centralizer** of $g \in G$ as

$$C(g) = \{x \in G \mid xg = xg\}.$$

Let $C_0(g)$ denote the identity component of $C(g)$, as defined in Exercise 7.7. Prove that $C_0(g)$ equals the union of all maximal tori of $G$ which contain $g$.

*Hint: If $x \in C_0(g)$, then $x$ belongs to a maximal torus of $C_0(g)$, which can be extended to a maximal torus of $G$.*

**Ex. 11.10.** If $a, b \in T$ are conjugate in $G$, prove that they are conjugate in $N(T)$. That is, if $g \cdot a \cdot g^{-1} = b$ for some $g \in G$, prove that $h \cdot a \cdot h^{-1} = b$ for some $h \in N(T)$.

*Hint: If $g \cdot a \cdot g^{-1} = b$, then $T$ and $g \cdot T \cdot g^{-1}$ are two maximal tori of $C_0(b)$, so one is a conjugate of the other inside $C_0(b)$. That is, there exists $x \in C_0(b)$ such that $xg \cdot T \cdot g^{-1}x^{-1} = T$. Now choose $h = xg$.*

**Ex. 11.11.** When we studied double covers in Section 7 of Chapter 8, we claimed:

$$sp(1) \cong so(3), \;\; sp(1) \times sp(1) \cong so(4), \;\; sp(2) \cong so(5), \;\; su(4) \cong so(6).$$

For each of these Lie algebra isomorphisms, show that the corresponding pair of Dynkin diagrams is identical. Show that no other pair of Dynkin diagrams of classical groups is identical, and thus that there are no other classical Lie algebra isomorphisms.

**Ex. 11.12.** Draw the root systems for the classical rank 2 groups: $SU(3)$, $SO(4)$, $SO(5)$, $Sp(2)$. The only other rank 2 root system is pictured in Figure 2.

**Ex. 11.13.** Prove that every rank 2 root system is one of the root systems from the previous exercise.

*Hint: The minimal angle, $\theta$, between any pair of roots must be $30°$, $45°$, $60°$, or $90°$. If $\alpha, \beta_1$ are roots which achieve this minimal angle, prove that for any integer $n$, there exists a root $\beta_n$ which forms an angle of $n\theta$ with $\alpha$, for example, $\beta_2 = -w_{\beta_1} \star \alpha$. Show $|\beta_{n_1}| = |\beta_{n_2}|$ if $n_1$ and $n_2$ are either both odd or both even.*

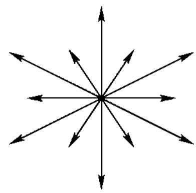


**Figure 2.** The root system of the exceptional group $G_2$.

**Ex. 11.14.** Prove Proposition 11.10, which says that the regular vectors of $G$ form an open dense subset of $\tau$.

**Ex. 11.15.** Let $(\tau, R)$ be a root system and let $\alpha, \beta \in R$. If $\angle(\alpha,\beta)$ is acute, prove that $\alpha - \beta \in R$. If $\angle(\alpha,\beta)$ is obtuse, prove that $\alpha + \beta \in R$.

*Hint: See the proof of Lemma* 11.31*. Note: From the dual root system of $G$, one can reconstruct its entire Lie algebra and bracket operation, which at least requires knowing which dual roots add to or subtract from which dual roots. This exercise give a glimpse of how such information can be obtained just from data about the angles between dual roots.*

**Ex. 11.16.** Prove that the Lie algebra of the center of $G$ equals

$$\mathfrak{z}(\mathfrak{g}) = \{A \in \mathfrak{g} \mid [A, X] = 0 \text{ for all } X \in \mathfrak{g}\},$$

which is called the **center** of $\mathfrak{g}$.

**Ex. 11.17.** For each $G \in \{SU(n), SO(2n), SO(2n+1), Sp(n)\}$, check that $W(G)$ acts transitively on the set of dual roots of a fixed length. That is, if $\hat{\alpha}, \hat{\beta}$ are dual roots with the same length, then there exists $w \in W(G)$ such that $w \star \hat{\alpha} = \hat{\beta}$.

**Ex. 11.18.** Let $\Delta = \{\alpha_1, ..., \alpha_t\}$ be a base of the root system $(\tau, R)$. Prove that for any $w \in W(\tau, R)$, $w \star \Delta = \{w \star \alpha_1, ..., w \star \alpha_t\}$ is also a base of the root system. It is also true that every base of the root system equals $w \star \Delta$ for some $w \in W(\tau, R)$.

**Ex. 11.19.** If Lemma 11.8 were false, so that, for example, $\hat{\alpha}_1 = \hat{\alpha}_2$, which is equivalent to $\alpha_1 = \alpha_2$, show that this would allow multiple ways for the 4-dimensional space $\mathfrak{l}_1 \oplus \mathfrak{l}_2$ to split into a pair of 2-dimensional $\mathrm{Ad}_T$-invariant spaces, for example:

$$\mathfrak{l}_1 \oplus \mathfrak{l}_2 = \mathrm{span}\{E_1 + E_2, F_1 + F_2\} \oplus \mathrm{span}\{E_1 - E_2, F_1 - F_2\}.$$

Thus, the decomposition of Theorem 11.1 would not be unique, and we would therefore have extra roots and dual roots corresponding to the extra possible $\mathrm{Ad}_T$-invariant decompositions of $\mathfrak{g}$.

**Ex. 11.20.** An element of $so(n)_{\mathbb{C}}$ has the form $X = X_1 + X_2\mathbf{i}$ for some $X_1, X_2 \in so(n)$. Interpret such an $X$ as an element of

$$so(n, \mathbb{C}) = \{A \in M_n(\mathbb{C}) \mid A + A^T = 0\} \subset M_n(\mathbb{C}).$$

Via this interpretation, show that the complexified Lie bracket operation in $so(n)_{\mathbb{C}}$ becomes identified with the following operation in $so(n, \mathbb{C})$: $[A, B]_{\mathbb{C}} = AB - BA$.

**Ex. 11.21.**

(1) Prove that every $X \in gl(n, \mathbb{C})$ can be uniquely expressed as $X = X_1 + X_2\mathbf{i}$ for $X_1, X_2 \in u(n)$. Further, $X \in sl(n, \mathbb{C})$ if and only if $X_1, X_2 \in su(n)$.

*Hint:* $X = \frac{X - X^*}{2} + \frac{X + X^*}{2\mathbf{i}}\mathbf{i}$.

(2) Use the above decomposition to identify $u(n)_{\mathbb{C}} \cong gl(n, \mathbb{C})$ and $su(n)_{\mathbb{C}} \cong sl(n, \mathbb{C})$. Show that the complex Lie bracket operations in $u(n)_{\mathbb{C}}$ and $su(n)_{\mathbb{C}}$ become identified with the operations in $gl(n, \mathbb{C})$ and $sl(n, \mathbb{C})$ defined as:

$$[A, B]_{\mathbb{C}} = AB - BA.$$

**Ex. 11.22.** Recall the injective function $\rho_n : M_n(\mathbb{C}) \to M_{2n}(\mathbb{R})$ from Chapter 2.

(1) Show that $\rho_n(SU(n))$ is a subgroup of $SO(2n)$ which is isomorphic to $SU(n)$.

(2) Show that $\rho_n(su(n))$ is a subalgebra of $so(2n)$ which is isomorphic to $su(n)$.

(3) Show that $\langle \rho_n(X), \rho_n(Y) \rangle = 2 \cdot \langle X, Y \rangle$ for all $X, Y \in M_n(\mathbb{C})$, so $\rho_n$ provides a non-orthogonal equivalence between the system of dual root of $su(n)$ and a subsystem of the system of dual roots of $so(2n)$.

**Ex. 11.23.** Prove that Table 2 is valid.

# Bibliography


1. A. Baker, *Matrix Groups: An Introduction to Lie Group Theory*, Springer, 2002.
2. I. Cornfeld, S. Fomin, Ya. Sinai, *Ergodic Theory*, Springer-Verlag, 1982.
3. M. Curtis, *Matrix Groups, Second Edition*, Springer, 1975, 1984.
4. G. Frobenius, Journal für die Reine und Angewandte Mathematik, 1878, Vol. 84, 1-63.
5. J. Gallian, *Contemporary Abstract Algebra*, Houghton Mifflin Co., 2002.
6. F. Goodman, *Algebra: Abstract and Concrete Stressing Symmetry*, Prentice Hall, 2003.
7. B. Hall, *Lie Groups, Lie Algebras, and Representations*, Springer, 2003.
8. F. R. Harvey, *Spinors and Calibrations*, Perspectives in Mathematics, Vol. 9, Academic Press, 1990.
9. S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, American Mathematical Society, 2001.
10. R. Howe, *Very basic Lie theory*, American Math. Monthly **90** (1983), 600-623; Correction, Amer. Math. Monthly **91** (1984), 247.
11. W. Rossmann, *Lie Groups: An Introduction Through Linear Groups*, Oxford Science Publications, 2002.
12. M. Spivak, *A Comprehensive Introduction to Differential Geometry, Volume 1*, Publish or Perish, 1979.
13. F. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, Springer-Verlag, 1983.
14. J. Weeks, *The Poincaré dodecahedral space and the mystery of the missing fluctuations*, Notices of the AMS **51** (2004), number 6, 610-619.

# Index

SELECTED PUBLISHED TITLES IN THIS SERIES

79 **Kristopher Tapp,** Matrix Groups for Undergraduates, Second Edition, 2016

78 **Gail S. Nelson,** A User-Friendly Introduction to Lebesgue Measure and Integration, 2015

77 **Wolfgang Kühnel,** Differential Geometry: Curves — Surfaces — Manifolds, Third Edition, 2015

76 **John Roe,** Winding Around, 2015

75 **Ida Kantor, Jiří Matoušek, and Robert Šámal,** Mathematics++, 2015

74 **Mohamed Elhamdadi and Sam Nelson,** Quandles, 2015

73 **Bruce M. Landman and Aaron Robertson,** Ramsey Theory on the Integers, Second Edition, 2014

72 **Mark Kot,** A First Course in the Calculus of Variations, 2014

71 **Joel Spencer,** Asymptopia, 2014

70 **Lasse Rempe-Gillen and Rebecca Waldecker,** Primality Testing for Beginners, 2014

69 **Mark Levi,** Classical Mechanics with Calculus of Variations and Optimal Control, 2014

68 **Samuel S. Wagstaff, Jr.,** The Joy of Factoring, 2013

67 **Emily H. Moore and Harriet S. Pollatsek,** Difference Sets, 2013

66 **Thomas Garrity, Richard Belshoff, Lynette Boos, Ryan Brown, Carl Lienert, David Murphy, Junalyn Navarra-Madsen, Pedro Poitevin, Shawn Robinson, Brian Snyder, and Caryn Werner,** Algebraic Geometry, 2013

65 **Victor H. Moll,** Numbers and Functions, 2012

64 **A. B. Sossinsky,** Geometries, 2012

63 **María Cristina Pereyra and Lesley A. Ward,** Harmonic Analysis, 2012

62 **Rebecca Weber,** Computability Theory, 2012

61 **Anthony Bonato and Richard J. Nowakowski,** The Game of Cops and Robbers on Graphs, 2011

60 **Richard Evan Schwartz,** Mostly Surfaces, 2011

59 **Pavel Etingof, Oleg Golberg, Sebastian Hensel, Tiankai Liu, Alex Schwendner, Dmitry Vaintrob, and Elena Yudovina,** Introduction to Representation Theory, 2011

58 **Álvaro Lozano-Robledo,** Elliptic Curves, Modular Forms, and Their L-functions, 2011

For a complete list of titles in this series, visit the
AMS Bookstore at **www.ams.org/bookstore/stmlseries/**.

Matrix groups touch an enormous spectrum of the mathematical arena. This textbook brings them into the undergraduate curriculum. It makes an excellent one-semester course for students familiar with linear and abstract algebra and prepares them for a graduate course on Lie groups.

*Matrix Groups for Undergraduates* is concrete and example-driven, with geometric motivation and rigorous proofs. The story begins and ends with the rotations of a globe. In between, the author combines rigor and intuition to describe the basic objects of Lie theory: Lie algebras, matrix exponentiation, Lie brackets, maximal tori, homogeneous spaces, and roots.

This second edition includes two new chapters that allow for an easier transition to the general theory of Lie groups.

**From reviews of the First Edition:**

*This book could be used as an excellent textbook for a one semester course at university and it will prepare students for a graduate course on Lie groups, Lie algebras, etc. . . . The book combines an intuitive style of writing with rigorous definitions and proofs, giving examples from fields of mathematics, physics, and other sciences where matrices are successfully applied. The book will surely be interesting and helpful for students in algebra and their teachers.*

—***European Mathematical Society Newsletters***

*This is an excellent, well-written textbook which is strongly recommended to a wide audience of readers interested in mathematics and its applications. The book is suitable for a one semester undergraduate lecture course in matrix groups, and would also be useful supplementary reading for more general group theory courses.*

—***MathSciNet* (or *Mathematical Reviews*)**

ISBN: 978-1-4704-2722-1

9 781470 427221

STML/79

For additional information and updates on this book, visit
**www.ams.org/bookpages/stml-79**