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 8.7).
- Movie graphics programmers use matrix groups for rotating and translating three-dimensional objects on a computer screen (see Section 3.6).
- 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 8.6). 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 3.2). 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 2×2 and some 4×4, 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, beginning with the cover artwork. 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 would like to thank Frank Morgan, Ed Burger, Tom Garrity, Phil Straffin, Wolfgang Ziller and Satyan Devadoss for sharing valuable suggestions. 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 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 and South 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 $SO(3) \cong S^2 \times S^1$ would help, allowing any rotation to be coded using only three real numbers – two which locate a point of S^2 and one angle which 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 ($a \cdot b = b \cdot a$) 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 (no element times $(5, 0)$ equals $(1, 1)$, which is the only possible

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 a more clever choice of multiplication operation is:

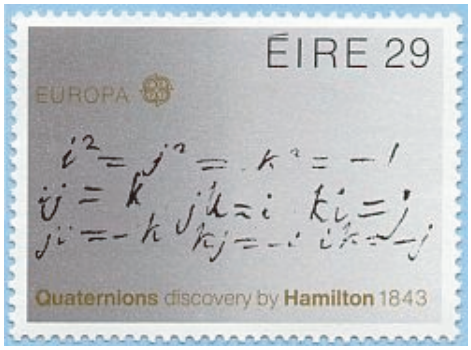
$$(a, b) \cdot (c, d) := (ac - bd, ad + bc).$$

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:

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

This sign convention can be remembered using Figure 2.

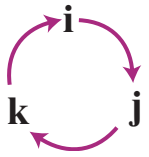


Figure 2. The quaternionic multiplication rule.

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 $\mathbf{k}$ or $\mathbf{j}$ 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 numbers 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 which 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, which 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 sub-field 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,

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

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:

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

Matrix multiplication is not generally commutative.

Denote a diagonal matrix as in this example:

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

The identity matrix is:

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

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

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

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:

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

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 3×3 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') $(A \cdot a) \cdot b = A \cdot (a \cdot b)$.

Do you see the difference? The net effect of multiplying A by a and then by b is to multiply A by ba in a left vector space, or by ab 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 $X \in \mathbb{K}^n$,*

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

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 $X, Y \in \mathbb{K}^n$,

$$R_A(aX + bY) = (aX + bY) \cdot A = a(X \cdot A) + b(Y \cdot A)$$
$$= a \cdot R_A(X) + b \cdot R_A(Y).$$

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 ith 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:*

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

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, $R_{BA} = R_A \circ R_B = \mathrm{id}$, which implies $BA = I$. Similarly, $R_{AB} = R_B \circ R_A = \mathrm{id}$, which implies $AB = I$. $\square$

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:

$$(c_1 v_1 + \cdots + c_n v_n) \mapsto (c_1, ..., c_n). \tag{1.7}$$

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 which seems 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 which 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 the vector $(c_1, ..., c_n)$. 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:

$$\begin{aligned} R_{gAg^{-1}}(X) = (c_1, ..., c_n) g A g^{-1} &= (c_1 v_1 + \cdots + c_n v_n) A g^{-1} \\ &= R_{g^{-1}}((c_1 v_1 + \cdots + c_n v_n) \cdot A). \end{aligned}$$

□

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 $\{e_1 g, ..., e_n g\}$.

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 counterclockwise rotation through an angle θ.

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 $GL_{n_1+n_2}(\mathbb{K})$ which 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 $\mathbb{R}^3$ which makes $\mathbb{R}^3$ into a real division algebra.

Ex. 1.17. Describe a subgroup of $GL_{n+1}(\mathbb{R})$ which 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 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 ≥ 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 θ 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 $X, Y \in \mathbb{C}^n$,

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

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

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

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

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 $h = f_{2n} \circ g_n : \mathbb{H}^n \to \mathbb{R}^{4n}$.

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

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* $\mathbb{K}$,

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

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

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

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

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) $\Longrightarrow$ (2) is obvious. (2) $\Longrightarrow$ (3) because the rows of A equal $\{R_A(e_1), ..., R_A(e_n)\}$. To see that (3) $\Longleftrightarrow$ (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) $\Longrightarrow$ (1). If the rows of A are orthonormal, then for all $X = (x_1, ..., x_2), Y = (y_1, ..., y_n) \in \mathbb{K}^n$,

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

□

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 $\mathbb{R}^{2n}$ which 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. 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.

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

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

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

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

$$\langle X, Y\rangle_{\mathbb{R}} = 1/2(|X + Y|^2_{\mathbb{R}} - |X|^2_{\mathbb{R}} - |Y|^2_{\mathbb{R}}).$$

So 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 $\rho_n(A) \in O(n)$, 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 θ. 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:

$$SO(2) = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \middle| \theta \in [0, 2\pi) \right\}, \tag{3.7}$$

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

$SO(2)$ 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 3-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 Ψ_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, $\det(A) = 1$ and $\begin{pmatrix} z_2 & -w_1 \\ -w_2 & z_1 \end{pmatrix} = A^{-1} = A^* = \begin{pmatrix} \overline{z}_1 & \overline{w}_2 \\ \overline{w}_1 & \overline{z}_2 \end{pmatrix}$, 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 $\mathbb{R}^n$ which fix the origin and 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 $A \in O(n)$ and $X, Y \in \mathbb{R}^n$,

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

which proves that R_A is an isometry.

Conversely, suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry for which $f(0) = 0$. For any $X \in \mathbb{R}^n$,

$$|f(X)| = \text{dist}(f(X), 0) = \text{dist}(f(X), f(0)) = \text{dist}(X, 0) = |X|,$$

which shows that f preserves norms. We showed in the proof of Proposition 3.11 that inner products are determined by norms, so f also preserves inner products; that is, for all $X, Y \in \mathbb{R}^n$,

$$\langle f(X), f(Y) \rangle = \langle X, Y \rangle.$$

Let A be the matrix whose ith 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 norms and 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. □

$O(n)$ is the group of isometries of $\mathbb{R}^n$ which fix the origin and which 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 (Chapter 1, Section 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 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 which fix the origin is isomorphic to $O(n)$. An isometry, f, that does not fix the origin is not linear, so cannot equal to R_A for any matrix A. In this case, let $V = f(0)$, so the function $X \mapsto f(X) - V$ is an isometry which 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 which 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

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

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 $\mathrm{Isom}(\mathbb{R}^n)$ is isomorphic to the following subgroup of $GL_{n+1}(\mathbb{R})$:

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

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

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

This is the group of isometries of $\mathbb{R}^n$ which 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 $\mathbb{R}^n$ which 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 $S^n \subset \mathbb{R}^{n+1}$ equals the group of isometries of $\mathbb{R}^{n+1}$ with no translational component, which is isomorphic to the orthogonal group:

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

In an abstract algebra course, you probably met 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 which 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*

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

are respectively called the "direct" and "indirect" symmetries 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 direct symmetries have index 1 (meaning all symmetries are direct) is illustrated in Figure 1.

Symmetry groups of subsets of $\mathbb{R}^2$ are useful for studying objects which are essentially 2-dimensional, like snowflakes and certain

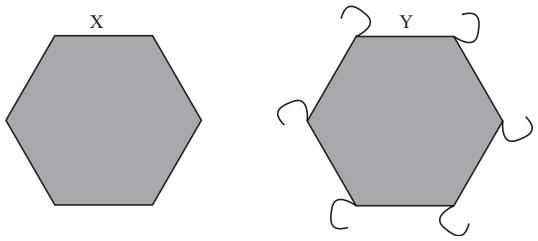


Figure 1. $\mathrm{Symm}(X) = D_6$, while $\mathrm{Symm}(Y) = \mathbb{Z}_6$.

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_4.

The *regular solids* provide examples of sets whose direct 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-

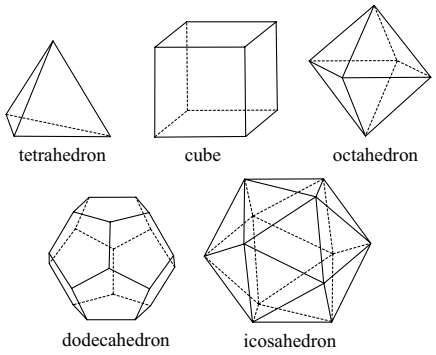


Figure 2. The five regular solids.

sidered to be sacred shapes, thought to represent fire, earth, air, the universe, and water. The fact that any other shape is "as symmetric" as one of these five (or is infinitely symmetric) enhances one's sense that the regular solids are of universal importance.

It turns out that A_4 is the direct symmetry group of a tetrahedron, S_4 is the direct symmetry group of a cube or an octahedron, and A_5 is the direct symmetry group of a dodecahedron or an icosahedron. See [**6**] for a complete calculation of these direct 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 direct symmetry group is S_4. What does a cube have 4 of which get permuted by its direct symmetries? It has 4 diagonals (lines connecting antipodal pairs of vertices). This observation is the starting point of the calculation of its direct 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 3.7)?

(2) Let $B = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \in SO(2)$. Assume that θ is not an integer multiple of π. 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 3.7).

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 direct 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 $\mathrm{Trans}(\mathbb{R}^n)$ is a normal subgroup of $\mathrm{Isom}(\mathbb{R}^n)$.

Ex. 3.11. Prove that the Affine group,

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

is a subgroup of $GL_{n+1}(\mathbb{K})$. Any $F \in \mathrm{Aff}_n(\mathbb{K})$ 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 lines in $\mathbb{K}^n$ to lines in $\mathbb{K}^n$. A 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 $\mathrm{Aff}_1(\mathbb{R})$ 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 $O(4)$ which is isomorphic to S_4 (S_4 = the group of permutations of a 4 elements set).

(3) Describe a subgroup, H, of $O(n)$ which 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 $A \in M_n(\mathbb{K})$ represent f in the basis $\mathcal{B}$. Prove that the following are equivalent:

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

(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.