Chapter 23

Rotation and Animation Using Quaternions

The previous chapter used complex analysis to further the study of minimal surfaces. Many applications of complex numbers to geometry can be generalized to the *quaternions*, an extended system in which the "imaginary part" of any number is a vector in $\mathbb{R}^3$. Although beyond the scope of this book, there is an extensive theory of surfaces defined by "quaternionic-holomorphic" curves [BFLPU]. Other uses include the visualization of fractals by means of iterated maps of quaternions extending the usual complex theory [DaKaSa, Wm2].

In this chapter, we describe the fundamental role that quaternions play in describing rotations of 3-dimensional space. It is a topic familiar to pure mathematicians (through the topology of Lie groups), but one that has recently assumed great popularity because of its use in computer graphics and video games. As a consequence, there is little in the chapter's text that cannot be found on the internet and elsewhere, though Notebook 23 contains a number of original animations based on the theory, hence the chapter's title.[1]

Applications are not restricted to merely viewing rotations; indeed many graphics interfaces already permit one to rotate a graphics object at the touch of a mouse. The purpose of this chapter is instead to help understand and develop the theory. For example, new surfaces can be constructed by rotating lines and curves in a nonstandard way, as in Figures 23.1 and 23.9.

During the course of the chapter, we give several descriptions of the group $\mathbf{SO}(3)$ of rotations in $\mathbb{R}^3$. All are based on representing a rotation by some sort of vector, and the fact that a rotation is uniquely specified by three parameters. In the first such description, in Section 23.1, an element $\mathbf{x}$ of $\mathbb{R}^3$ is converted (via the

[1] Added in proof: For a more extensive treatment of some of the topics in this chapter, the editors recommend the book by A.J. Hanson "Visualizing Quaternions," Elsevier, 2006.

vector cross product) into a skew-symmetric matrix A, and then exponentiated. This gives rise to a neat expression for a rotation of a given angle about a given axis, namely Theorem 23.4, whose proof is completed using quaternions.

After describing the basic operations on quaternions in Section 23.2, we define the celebrated mapping from the group $\mathscr{U}$ of unit quaternions to the group $\mathbf{SO}(3)$ of rotations in $\mathbb{R}^3$. This has a built-in ambiguity whereby plus and minus a quaternion describe the same rotation, and some of the more interesting graphics generated by Notebook 23 show tell-tale signs of this phenomenon, also present in complex analysis. The construction converts circles into figure eight curves and generates self-intersecting surfaces like that in Figure 23.1, which occur in the theory of Riemann surfaces [AhSa]. It all relates to the topological picture of $\mathbf{SO}(3)$ as a closed "ball" in $\mathbb{R}^3$ for which antipodal points of the boundary are identified, as explained in Section 23.4.

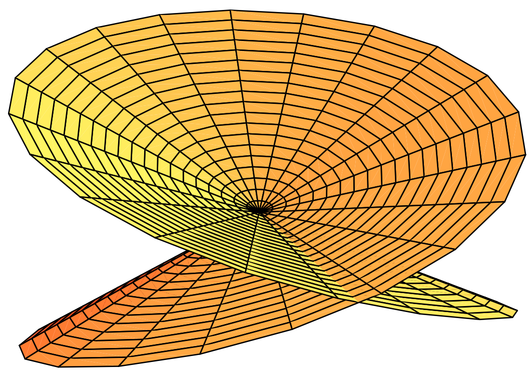

Figure 23.1: A ruled surface generated by a rotation curve

Rotation curves, the subject of Section 23.5, are simply curves whose trace lies in $\mathbf{SO}(3)$, rather than in $\mathbb{R}^3$. In practical terms, one can think of such a curve as a "static fairground ride" in which the rider is strapped inside a sphere which rotates about its center, with varying axis, and the parameter t is time. One can also associate a rotation curve to a more dynamic rollercoaster-type ride using the Frenet frame construction of Chapter 7. Section 23.7 briefly discusses this and other topics. In the intervening Section 23.6, we explain how Euler angles are used to decompose any rotation in space into the product of rotations about coordinate axes.

There is a sense in which this chapter is a bridge between past and future topics. The linear algebra reaches back to the start of the book, while Section 23.5 relates to homotopy and the theory of space curves in Chapters 6 and 7. On the other hand, the study of rotations and quaternions provides simple examples of differentiable manifolds and related techniques that we shall meet in the next chapter.

23.1 Orthogonal Matrices

As explained at the outset in Section 1.1, one can represent a linear mapping $\mathbb{R}^n \to \mathbb{R}^m$ by a matrix of size $m \times n$. In this chapter, we shall be mainly concerned with the case $m = n = 3$, but to begin with we impose only the restriction $m = n$. This being the case, one can compose two linear mappings A, B to make a third, denoted $A \circ B$, defined by

$$(A \circ B)(\mathbf{v}) = A(B(\mathbf{v})), \qquad \mathbf{v} \in \mathbb{R}^n. \tag{23.1}$$

If one regards A, B as matrices rather than mappings, then it is obvious from (23.1) that their composition $A \circ B$ corresponds to the matrix product AB. One only has to remember that it is actually B (or, more generally, the *second* factor) that is applied *first* to a column vector in $\mathbb{R}^n$.

By its very definition, composition of mappings $A, B, C \colon \mathbb{R}^n \to \mathbb{R}^n$ satisfies the associative law

$$A \circ (B \circ C) = (A \circ B) \circ C, \tag{23.2}$$

which forms part of the definition of a group (see page 131). The same is of course true of matrix multiplication, and in Section 5.1 we defined various groups by imposing restrictions on the linear transformations or matrices under consideration. In this chapter, the emphasis will be on matrices, and we begin with

Definition 23.1. An $n \times n$ matrix A is said to be ***orthogonal*** *if*

$$A^T A = I_n, \tag{23.3}$$

where I_n is the identity matrix of order n.

Recall that A^T denotes the transpose of A, obtained by swapping its rows and columns. Equation (23.3) therefore asserts that the *columns* of A form an orthonormal basis of $\mathbb{R}^n$. For this reason, it would be more logical to speak of an "orthonormal" matrix, though one never does.

Let $\mathbf{O}(n)$ denote the set of all orthogonal $n \times n$ matrices. Provided we represent a linear transformation with respect to a fixed orthonormal basis, $\mathbf{O}(n)$ coincides with the group $\mathsf{orthogonal}(\mathbb{R}^n)$, via equation (5.2). The fact that $\mathbf{O}(n)$ is a group can be seen directly by observing that if $A, B \in \mathbf{O}(n)$, then

$$(AB)^T(AB) = B^T A^T A B = B^T I_n B = I_n.$$

Moreover, the inverse of an orthogonal martrix is given by

$$A^{-1} = A^T,$$

from which it follows that $AA^T = I_n$, or that the *rows* of A also form an orthonormal basis.

Lemma 5.2 tells us that the determinant of any orthogonal matrix is either 1 or -1. Since $\det(AB) = (\det A)(\det B)$, it follows that the set

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

is a *subgroup* of $\mathbf{O}(n)$; that is, a group in its own right with the same operation of matrix multiplication. The letter $\mathbf{S}$ stands for "special," and $\mathbf{SO}(n)$ corresponds to the group of *orientation-preserving* orthogonal transformations, or *rotations*, as they were called on page 128.

We now specialize to the case $n = 3$ for which there is already a clear notion of "rotation," meaning *a transformation leaving fixed the points on some straight line axis, and rotating any other point in a plane perpendicular to the axis*. An example of such a rotation is the transformation represented by the matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{23.4}$$

where $0 \leqslant \theta < 2\pi$. This mapping leaves fixed the z-axis, but acts as a rotation by an angle θ counterclockwise in the xy-plane.

The following well-known result implies that any element of $\mathbf{SO}(3)$ is equivalent to some R_θ after a change of orthonormal basis.

Lemma 23.2. *Given a* 3×3 *matrix* $A \in \mathbf{SO}(3)$, *there exists* $P \in \mathbf{SO}(3)$ *and* $\theta \in [0, 2\pi)$ *such that*

$$P^{-1}AP = R_\theta.$$

Proof. The first step is to show that 1 is an eigenvalue of A. This is true because

$$\begin{aligned} \det(A - I_3) \;=\; \det(A - AA^T) \;&=\; \det\big(A(I_3 - A)^T\big) \\ &=\; (\det A)\det\big((I_3 - A)^T\big) \\ &=\; -\det(A - I_3) \end{aligned}$$

is zero. Since $A - I_3$ is singular, there exists a nonzero column vector $\mathbf{w}$ such that $A\mathbf{w} = \mathbf{w}$, and it is this eigenvector that determines the axis of rotation. Extend the unit vector $\mathbf{w}_3 = \mathbf{w}/\|\mathbf{w}\|$ to an orthonormal basis $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ such that $\mathbf{w}_1 \times \mathbf{w}_2 = \mathbf{w}_3$. Being a positively oriented *orthogonal transformation*, A must restrict to a rotation in the plane generated by $\mathbf{w}_1, \mathbf{w}_2$, and there exists θ such that

$$\begin{cases} A\mathbf{w}_1 \;=\; \mathbf{w}_1\cos\theta + \mathbf{w}_2\sin\theta \\ A\mathbf{w}_2 \;=\, -\mathbf{w}_1\sin\theta + \mathbf{w}_2\cos\theta. \end{cases} \tag{23.5}$$

Now let P denote the invertible matrix whose columns are the vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$. By definition of matrix multiplication, the columns of AP are $A\mathbf{w}_1,\ A\mathbf{w}_2,\ A\mathbf{w}_3$, and (23.5) now implies that $AP = PR_\theta$. The result follows. ▮

Lemma 23.2 can be expressed by saying that A is ***conjugate*** to R_θ in the group $\mathbf{SO}(n)$. The two rotations are equivalent, if we are willing to change our frame of reference by pointing the z-axis in the direction of the eigenvector $\mathbf{w}$. Combining (23.6) with equation (13.18) on page 398 gives

$$\operatorname{tr} A = \operatorname{tr}(PR_\theta P^{-1}) = \operatorname{tr}(R_\theta) = 2\cos\theta + 1. \tag{23.6}$$

This means that the unsigned angle $\pm\theta$ can conveniently be recovered from the matrix A. To see how to recover the axis of rotation, see Exercise 4.

Later in this chapter, we shall be considering *curves* in the group $\mathbf{SO}(3)$. Such a curve is a mapping $\gamma\colon (a,b) \to \mathbb{R}^{3,3}$ where $(a,b) \subseteq \mathbb{R}$ and $\mathbb{R}^{3,3}$ is the space of 3×3 matrices, and the problem is to find conditions that ensure that the image of γ lies in $\mathbf{SO}(3)$. One method is to use a power series

$$\gamma(t) = I + A_1 t + A_2 t^2 + \cdots \tag{23.7}$$

where $I = I_3$, and seek appropriate conditions on the coefficients. For example, to determine A_1, we need only consider the terms up to order one, and set

$$\begin{aligned} I = \gamma(t)\gamma(t)^T &= (I + A_1 t + \cdots)(I + A_1 t + \cdots)^T \\ &= I + (A_1 + A_1^T)t + \cdots \end{aligned}$$

It follows that A_1 must be a *skew-symmetric* matrix, meaning $A_1{}^T = -A_1$.

Having made the crucial step, it is in fact possible to solve (23.7) using the concept of exponentiation. The ***exponential*** of a square matrix A, written e^A or $\exp A$, is defined by the usual infinite sum

$$\exp A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k, \tag{23.8}$$

with the convention that $A^0 = I$. This operation is well known to share some of the usual properties of exponentiation of real numbers. In particular, (23.8) converges, and if A, B are two square matrixes for which $AB = BA$ then

$$\exp(A+B) = (\exp A)(\exp B).$$

If $\mathbf{0}$ denotes the zero matrix then obviously $\exp(\mathbf{0}) = I$. Thus,

$$\exp(-A) = (\exp A)^{-1},$$

and the matrix (23.8) is always invertible.

Any property valid for the powers of a matrix will extend to the operation of exponentiation. For example, $(A^n)^T = (A^T)^n$ for all $n \geqslant 0$, and so $\exp(A^T) = (\exp A)^T$. A more interesting formula is

$$\det(\exp A) = \exp(\operatorname{tr} A). \tag{23.9}$$

This is immediate if A is *diagonal* with entries $\lambda_1, \dots, \lambda_n$, for then e^A is diagonal with entries $e^{\lambda_1}, \dots e^{\lambda_n}$, and (23.9) becomes the identity

$$\prod_{i=1}^{n} e^{\lambda_i} = e^{\lambda_1 + \cdots + \lambda_n}.$$

Having seen this, the validity of (23.9) can easily be extended to the case in which A is *diagonalizable* (over the complex numbers). For then there exists a complex invertible matrix P and a complex diagonal matrix D for which $A = PDP^{-1}$, and

$$\begin{aligned} \exp A \;=\; \exp(PDP^{-1}) \;&=\; \sum_{k=0}^{\infty} \frac{1}{k!}(PDP^{-1})^k \\ &=\; \sum_{k=0}^{\infty} \frac{1}{k!} PD^k P^{-1} \;=\; P(\exp D)P^{-1} \end{aligned}$$

The result (23.9) now follows from the properties (13.18) already used for (23.6).

It is a general principle that, to prove a matrix equation like (23.9) (in which both sides are continuous matrix functions), it suffices to assume that A is diagonalizable. This is because an arbitrary square matrix can be approximated by a sequence of diagonalizable matrices (for example, ones that have distinct complex eigenvalues). In this way (23.9) follows in general; see [Warner] for a fuller discussion of the exponentiation of matrices. The same principle can be used to prove the *Cayley–Hamilton Theorem* (mentioned on page 402), namely that any $n \times n$ matrix A satisfies its characteristic polynomial

$$c_A(x) = \det(xI_n - A).$$

This theorem implies that the nth power A^n is always some linear combination of $I_n, A, A^2, \dots, A^{n-1}$. The same is therefore true of $\exp A$, and a vivid illustration of this fact is provided by Theorem 23.4 below.

Lemma 23.3. *If A is a $n \times n$ skew-symmetric matrix, then $\exp A$ is an orthogonal matrix with determinant 1.*

Proof. We have

$$\exp(A^T A) = \exp(A^T) \exp A = \exp(-A) \exp A = (\exp A)^{-1} \exp A = I_n.$$

Given that $\operatorname{tr} A = 0$, the statement about $\det A = 1$ follows from (13.18). ▮

The importance of this lemma can be assessed by the fact that it is related to at least two topics studied in later chapters.[2]

Any skew-symmetric 3×3 matrix is uniquely determined by a column vector

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$$

by means of the formula

$$A[\mathbf{x}] = \begin{pmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{pmatrix}. \tag{23.10}$$

We need not have used coordinates, since (23.10) is equivalent to stating that

$$A[\mathbf{x}]\mathbf{w} = \mathbf{x} \times \mathbf{w} \tag{23.11}$$

for all $\mathbf{w} \in \mathbb{R}^3$. To understand this, consider

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{23.12}$$

and observe that the ith row of $A[\mathbf{x}]$ is (the transpose of) the vector cross product $\mathbf{e}_i \times \mathbf{x}$. Thus, the ith entry of the column vector $A[\mathbf{x}]\mathbf{w}$ equals

$$(\mathbf{e}_i \times \mathbf{x}) \cdot \mathbf{w} = \mathbf{e}_i \cdot (\mathbf{x} \times \mathbf{w}),$$

explaining the appearance of the cross product of $\mathbf{x}$ with $\mathbf{w}$.

The next result is a basic example for the mathematical theory of Lie groups [DuKo], and is also well known in the context of computer vision [Faug].

Theorem 23.4. (i) *Let* $\mathbf{x}$ *be a nonzero vector in* $\mathbb{R}^3$ *and set* $\ell = \|\mathbf{x}\|$. *Then*

$$\exp A[\mathbf{x}] = I + \frac{\sin \ell}{\ell} A[\mathbf{x}] + \frac{1 - \cos \ell}{\ell^2} A[\mathbf{x}]^2. \tag{23.13}$$

(ii) *any nonidentity* $P \in \mathbf{SO}(3)$ *can be written in this form; indeed* (23.13) *represents a rotation with axis parallel to* $\mathbf{x}$ *and angle* ℓ.

We choose to postpone the proof of (ii), which demonstrates the simplicity with which rotations can be represented via linear algebra. It is accomplished by Corollary 23.14.

---

[2] The set of skew-symmetric $n \times n$ matrices forms what is called a *Lie algebra* when one defines $[A, B] = AB - BA$ (see Exercise 5 and page 832). There exists a means of exponentiating any finite-dimensional Lie algebra to obtain a *Lie group*, thus generalizing Lemma 23.3 [Warner]. There is a more general notion of exponentiation in Riemannian geometry, whereby the tangent space of a surface acts to model the surface itself (see page 885).

Proof of (i). Let $\ell = \|\mathbf{x}\|$, and set $\mathbf{x}_1 = \mathbf{x}/\ell$. Choose a unit vector $\mathbf{x}_2$ orthogonal to $\mathbf{x}$ and let $\mathbf{x}_3 = \mathbf{x}_1 \times \mathbf{x}_2$. With respect to the basis $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, the linear transformation (23.11) associated to $A[\mathbf{x}]$ has matrix

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\ell \\ 0 & \ell & 0 \end{pmatrix}. \tag{23.14}$$

There is a similarity with (23.4), explained by the fact that (23.14) equals $A[\ell\mathbf{e}_1]$, and $A[\mathbf{x}]$ acts as ℓ times the operator J in the plane generated by $\mathbf{x}_2$ and $\mathbf{x}_3$. Both $A[\ell\mathbf{e}_1]$ and $A[\mathbf{x}]$ have eigenvalues $0, i\ell, -i\ell$. Indeed, $A[\ell\mathbf{e}_1] = PDP^{-1}$ where

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & i & 1 \\ 0 & 1 & i \end{pmatrix}, \qquad D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & i\ell & 0 \\ 0 & 0 & -i\ell \end{pmatrix}.$$

The matrix D satisfies the equation

$$e^x = 1 + \frac{\sin\ell}{\ell}x + \frac{1-\cos\ell}{\ell^2}x^2$$

(with D in place of x and I_3 in place of the first 1) because each of its diagonal entries does. The same equation is satisfied by $A[\mathbf{x}]$ because

$$\begin{aligned} \exp(PDP^{-1}) &= P(\exp D)P^{-1} \\ &= I_3 + \frac{\sin\ell}{\ell}PDP^{-1} + \frac{1-\cos\ell}{\ell^2}(PDP^{-1})^2. \ \blacksquare \end{aligned}$$

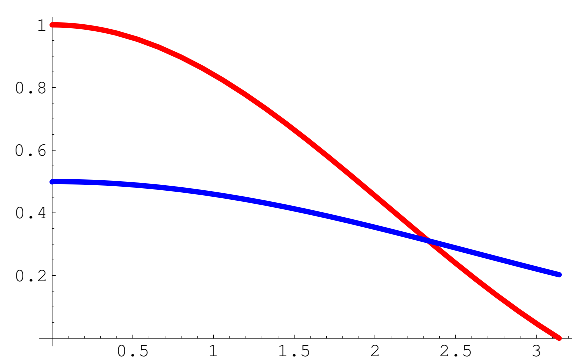

Figure 23.2: The functions $\dfrac{\sin\ell}{\ell}$ and $\dfrac{1-\cos\ell}{\ell^2}$

Figure 23.2 plots the functions occurring in (23.13) over the range $0 \leqslant \ell \leqslant \pi$ that will interest us in Section 23.4. Both tend to a limit as $\ell \to 0$, so we may allow $\mathbf{x} = \mathbf{0}$ in (23.13), as this is consistent with the formula $\exp\mathbf{0} = I$.

We shall return to the theory underlying Theorem 23.4 after the introduction of quaternions.

23.2 Quaternion Algebra

At one level, a quaternion is merely a vector with four components — that is an element of 4-space — and we explore this aspect first. By analogy to complex numbers, one denotes an individual quaternion by

$$\mathbf{q} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}, \qquad a, b, c, d \in \mathbb{R}, \tag{23.15}$$

and the set of quaternions by $\mathbb{H}$. Since $\mathbf{q}$ is determined by the vector (a, b, c, d), the set $\mathbb{H}$ can be identified with $\mathbb{R}^4$ as a vector space. But the real beauty of $\mathbb{H}$ derives from the special way in which its elements can be multiplied.

The multiplication of quaternions is based on the "fundamental formula"

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

that Hamilton[3] carved into Brougham bridge in 1843. It satisfies the associative rule (analogous to (23.2) without the $\circ$), which asserts that there is no ambiguity in computing a product of three or more quaternions written in a fixed order (this is Corollary 23.6 overleaf). For example,

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

Indeed, it is frequently convenient to regard $\mathbf{i}, \mathbf{j}, \mathbf{k}$ as equal to the respective vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ in (23.12). Their multiplication is then derived from the vector cross product in $\mathbb{R}^3$.

Now that we have defined quaternion multiplication, one can appreciate that the four coordinates a, b, c, d on the right-hand side of (23.15) are not equivalent. Indeed, $\mathbb{R}^4$ is split up into $\mathbb{R}$ and $\mathbb{R}^3$, and we can distinguish the ***real part***

$$\mathfrak{Re}\,\mathbf{q} = a,$$

and the ***imaginary part***

$$\mathfrak{Im}\,\mathbf{q} = b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$$

of any quaternion. Unlike for complex numbers (in which the imaginary part of $z = x + iy$ is the real number y), the imaginary part of $\mathbf{q}$ is effectively a *vector* in $\mathbb{R}^3$. One says that $\mathbf{q}$ is an ***imaginary quaternion*** if $\mathfrak{Re}\,\mathbf{q} = 0$.

We shall often write

$$\mathbf{q} = a + \mathbf{v}, \tag{23.16}$$

with the understanding that $\mathbf{v} = \mathfrak{Im}\,\mathbf{q} \in \mathbb{R}^3$.

[3] See page 191.

Lemma 23.5. *If* $\mathbf{q}_1 = a_1 + \mathbf{v}_1$ *and* $\mathbf{q}_2 = a_2 + \mathbf{v}_2$*, then*

$$\begin{aligned} \mathfrak{Re}(\mathbf{q}_1\mathbf{q}_2) &= a_1a_2 - \mathbf{v}_1 \cdot \mathbf{v}_2, \\ \mathfrak{Im}(\mathbf{q}_1\mathbf{q}_2) &= a_1\mathbf{v}_2 + a_2\mathbf{v}_1 + \mathbf{v}_1 \times \mathbf{v}_2. \end{aligned} \tag{23.17}$$

In particular, if $\mathbf{q}_1 = \mathbf{v}_1$ *and* $\mathbf{q}_2 = \mathbf{v}_2$ *are imaginary quaternions, then*

$$\mathbf{q}_1\mathbf{q}_2 = -\mathbf{v}_1 \cdot \mathbf{v}_2 + \mathbf{v}_1 \times \mathbf{v}_2. \tag{23.18}$$

Proof. This is just a matter of expanding brackets:

$$\begin{aligned} \mathbf{q}_1\mathbf{q}_2 &= (a_1 + \mathbf{v}_1)(a_2 + \mathbf{v}_2) \\ &= (a_1a_2 - \mathbf{v}_1 \cdot \mathbf{v}_2) + (a_1\mathbf{v}_2 + a_2\mathbf{v}_1 + \mathbf{v}_1 \times \mathbf{v}_2). \end{aligned}$$

The last statement follows from setting $a_1 = 0 = a_2$. ∎

The next result is verified in Notebook 23.

Corollary 23.6. *Multiplication of quaternions is associative:*

$$(\mathbf{q}_1\mathbf{q}_2)\mathbf{q}_3 = \mathbf{q}_1(\mathbf{q}_2\mathbf{q}_3),$$

for all $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3 \in \mathbb{H}$.

Of course, quaternion multiplication is not commutative, by its very definition (but see Exercise 2).

The following will help us emphasize similarities between $\mathbb{H}$ and $\mathbb{C}$.

Definition 23.7. *Let* $\mathbf{q} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} = a + \mathbf{v} \in \mathbb{H}$.

(i) *The* ***modulus*** *or* ***norm*** *of* $\mathbf{q}$*, written* $|\mathbf{q}|$*, is given by*

$$|\mathbf{q}|^2 = a^2 + b^2 + c^2 + d^2 = a^2 + \|\mathbf{v}\|^2.$$

Moreover, $\mathbf{q}$ *is called a* ***unit quaternion*** *if* $|\mathbf{q}| = 1$.

(ii) *The* ***conjugate*** *of* $\mathbf{q}$*, written* $\overline{\mathbf{q}}$*, is given by*

$$\overline{\mathbf{q}} = a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k} = a - \mathbf{v}.$$

Thus, the norm of $\mathbf{q}$ is simply its norm as a vector in $\mathbb{R}^4$, as defined on page 3. Because of the analogy with complex numbers (and anticipating Lemma 23.8), we indicate the norm with a single vertical line, rather than two. For consistency, we shall also indicate the norm of a vector $\mathbf{v} \in \mathbb{R}^3$ by $|\mathbf{v}|$, rather than $\|\mathbf{v}\|$, because in this chapter we shall usually be thinking of such a vector as the imaginary quaternion $\mathbf{q} = 0 + \mathbf{v}$.

The conjugate of a quaternion $\mathbf{q}$ is formed by reversing the sign of its imaginary component. We shall use the expression $\mathfrak{Im}\,\mathbb{H}$ to indicate the space

$$\{\mathbf{q} \in \mathbb{H} \mid \overline{\mathbf{q}} = -\mathbf{q}\}$$

of imaginary quaternions. It is effectively $\mathbb{R}^3$, but using the definitions above, we can multiply two elements of $\mathfrak{Im}\,\mathbb{H}$ according to the rule (23.18). In particular, if $\mathbf{q} = \mathbf{v}$ is an imaginary quaternion, then

$$\mathbf{q}^2 = \mathbf{v}^2 = -\mathbf{v} \cdot \mathbf{v} = -|\mathbf{v}|^2 \tag{23.19}$$

is a *real* number.

The key properties underlying Definition 23.7 are encapsulated in

Lemma 23.8. *Let* $\mathbf{q}, \mathbf{q}_1, \mathbf{q}_2 \in \mathbb{H}$. *Then*

(i) $\mathbf{q}\overline{\mathbf{q}} = |\mathbf{q}|^2$;

(ii) $\overline{\mathbf{q}_1\mathbf{q}_2} = \overline{\mathbf{q}}_2\,\overline{\mathbf{q}}_1$;

(iii) $|\mathbf{q}_1\mathbf{q}_2| = |\mathbf{q}_1||\mathbf{q}_2|$.

Proof. Part (i) follows from (23.19), given that $(a + \mathbf{v})(a - \mathbf{v}) = a^2 - \mathbf{v}^2$. Part (ii) is an immediate consequence of equation (23.17), and the fact that $-\mathbf{v}_1 \times \mathbf{v}_2 = \mathbf{v}_2 \times \mathbf{v}_1$. Given (i),

$$\begin{aligned} |\mathbf{q}_1\mathbf{q}_2|^2 &= \mathbf{q}_1\mathbf{q}_2\,\overline{\mathbf{q}_1\mathbf{q}_2} \\ &= \mathbf{q}_1\mathbf{q}_2\,\overline{\mathbf{q}}_2\,\overline{\mathbf{q}}_1 = \mathbf{q}_1|\mathbf{q}_2|^2\overline{\mathbf{q}}_1 = |\mathbf{q}_1|^2|\mathbf{q}_2|^2. \quad \blacksquare \end{aligned}$$

Part (i) of this lemma has an important consequence. It exhibits explicitly the inverse of a nonzero quaternion, namely

$$\mathbf{q}^{-1} = \frac{1}{|\mathbf{q}|}\overline{\mathbf{q}}, \qquad \mathbf{q} \neq 0.$$

This means that $\mathbb{H}$ is a ***division ring*** or ***skew-field***, an algebraic structure with all the properties of a field except commutativity [Her, CoSm].

At this point, the multiplication rules allow one to regard a complex number as a quaternion for which $c = d = 0$. However, this is to miss the point, since we could have equally well chosen $\mathbf{j}$ or $\mathbf{k}$ to play the role of the symbol $i = \sqrt{-1}$ used to define complex numbers. Indeed, if $\mathbf{v} \in \mathfrak{Im}\,\mathbb{H}$ is any fixed ***unit imaginary quaternion***, we can identify the real 2-dimensional subspace

$$\langle\, 1,\ \mathbf{v}\,\rangle = \{a + b\mathbf{v} \mid a, b \in \mathbb{R}\} \tag{23.20}$$

with $\mathbb{C}$ by means of the mapping

$$a + ib \mapsto a + b\mathbf{v}.$$

The multiplications (of $\mathbb{C}$ on the one hand and $\mathfrak{Im}\,\mathbb{H}$ on the other) are consistent, because

$$\begin{aligned}(a_1+b_1\mathbf{v})(a_2+b_2\mathbf{v}) &= a_1a_2+(a_1b_2+b_1a_2)\mathbf{v}+b_1b_2\mathbf{v}^2\\ &= (a_1a_2-b_1b_2)+(a_1b_2+b_1a_2)\mathbf{v}.\end{aligned}$$

The set of unit imaginary quaternions is parametrized by the unit sphere $S^2(1)$; in conclusion, there is a whole sphere's worth of complex planes incorporated into $\mathbb{H}$! This multiple way in which $\mathbb{H}$ extends $\mathbb{C}$ is what is understood by the term "hypercomplex geometry" when applied in relation to the quaternions.

Writing out Lemma 23.8(iii) (squared) in full gives the identity

$$(a_1{}^2+b_1{}^2+c_1{}^2+d_1{}^2)(a_2{}^2+b_2{}^2+c_2{}^2+d_2{}^2)=a_3{}^2+b_3{}^2+c_3{}^2+d_3{}^2, \tag{23.21}$$

where (copying from Notebook 23),

$$\begin{aligned}a_3 &= a_1a_2-b_1b_2-c_1c_2-d_1d_2,\\ b_3 &= a_2b_1+a_1b_2-c_2d_1+c_1d_2,\\ c_3 &= a_2c_1+a_1c_2+b_2d_1-b_1d_2,\\ d_3 &= -b_2c_1+b_1c_2+a_2d_1+a_1d_2.\end{aligned}$$

Now consider what happens when the components of $\mathbf{q}_1$ and $\mathbf{q}_2$ are all integers. We see that the product of two *sums of four squared integers* is itself a *sum of four squared integers*. This fact is the crucial first step in proving the following result of Lagrange.

Theorem 23.9. Any *positive integer n can be written as the sum*

$$n=a^2+b^2+c^2+d^2,\qquad a,b,c,d\in\mathbb{N} \tag{23.22}$$

of four squares of nonnegative integers.

In the light of (23.21), it is sufficient to prove the result when n is a prime number, and we refer the reader to any standard text on elementary number theory, such as [Baker].

The representation (23.22) is in general far from unique. For example, it is verified in Notebook 23 that if $n=100$ then the solutions (a,b,c,d) with $a\geqslant b\geqslant c\geqslant d$ are

$$(5,5,5,5),\ (7,5,5,1),\ (7,7,1,1),\ (8,4,4,2),\ (8,6,0,0),\ (9,3,3,1),\ (10,0,0,0).$$

Three of them are implicit in the quaternionic factorization

$$(1+i+j+k)(3-4j)=7+7i-j-k.$$

23.3 Unit Quaternions and Rotations

We have seen (Corollary 23.6) that quaternions obey the associative rule. The set $\mathbb{H}$ is not itself a group because the quaternion 0 has no multiplicative inverse, though the set $\mathbb{H} \setminus \{0\}$ *is* a group under multiplication. So is the set

$$\mathscr{U} = \{\mathbf{q} \in \mathbb{H} \mid |\mathbf{q}| = 1\}$$

of unit quaternions. This follows from Lemma 23.8 since if $\mathbf{q}_1, \mathbf{q}_2 \in \mathscr{U}$ then

$$|\mathbf{q}_1\mathbf{q}_2| = |\mathbf{q}_1||\mathbf{q}_2| = 1,$$

and $\mathscr{U}$ is closed under products. Also, if $\mathbf{q} \in \mathscr{U}$ then

$$\mathbf{q}^{-1} = \overline{\mathbf{q}} \in \mathscr{U},$$

showing that $\mathscr{U}$ is closed under taking inverses, and the rules on page 131 are satisfied. Readers who are not familiar with $\mathscr{U}$ may nonetheless be acquainted with its finite subgroup Q_8 of eight elements (see Exercise 3).

Since a unit quaternion has the form

$$\mathbf{q} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}, \qquad \text{with} \qquad a^2 + b^2 + c^2 + d^2 = 1,$$

$\mathscr{U}$ can be identified with the set

$$\big\{(a,b,c,d) \in \mathbb{R}^4 : a^2 + b^2 + c^2 + d^2 = 1\big\}. \tag{23.23}$$

This is none other than the hypersphere $S^3(1)$, consisting of all the 4-vectors a unit distance from the origin (see similar definitions on pages 236 and 662).

The aim of this section is to show that $\mathscr{U}$ is intimately related to the group $\mathbf{SO}(3)$ of rotations in $\mathbb{R}^3$. The starting point is

Lemma 23.10. *Let $\mathbf{q} \in \mathscr{U}$. If $\mathbf{w} \in \mathfrak{Im}\,\mathbb{H}$ then $\mathbf{q}\mathbf{w}\overline{\mathbf{q}}$ is an imaginary quaternion with the same norm as $\mathbf{w}$. The resulting mapping*

$$\mathbf{w} \mapsto \mathbf{q}\mathbf{w}\overline{\mathbf{q}} \tag{23.24}$$

is an orthogonal transformation of $\mathbb{R}^3$.

Proof. By hypothesis, $\overline{\mathbf{w}} = -\mathbf{w}$ and $\mathbf{w}^2 = -|\mathbf{w}|^2$. Let $\mathbf{p} = \mathbf{q}\mathbf{w}\overline{\mathbf{q}}$. The property of conjugation implies that

$$\overline{\mathbf{p}} = \overline{\overline{\mathbf{q}}}\,\overline{\mathbf{w}}\,\overline{\mathbf{q}} = \mathbf{q}(-\mathbf{w})\overline{\mathbf{q}} = -\overline{\mathbf{p}}.$$

It follows that $\mathbf{p}$ is also an imaginary quaternion; we compute

$$\begin{aligned} |\mathbf{p}|^2 \;=\; \mathbf{p}\overline{\mathbf{p}} &= (\mathbf{q}\mathbf{w}\overline{\mathbf{q}})(-\mathbf{q}\mathbf{w}\overline{\mathbf{q}}) \\ &= \mathbf{q}(-\mathbf{w}^2)\overline{\mathbf{q}} \\ &= \mathbf{q}|\mathbf{w}|^2\overline{\mathbf{q}} \;=\; |\mathbf{w}|^2. \end{aligned} \tag{23.25}$$

The mapping $\mathbf{w} \mapsto \mathbf{p}$ is certainly linear, and the last statement now follows from the proof of Theorem 5.6 on page 132. ∎

We denote the transformation (23.24) by $R[\mathbf{q}]$. Thus, if $\mathbf{q} = a + \mathbf{v} \in \mathscr{U}$, then

$$
\begin{aligned}
R[\mathbf{q}]\mathbf{w} &= (a+\mathbf{v})\mathbf{w}(a-\mathbf{v}) \\
&= a^2\mathbf{w} + a(\mathbf{v}\mathbf{w} - \mathbf{w}\mathbf{v}) - \mathbf{v}\mathbf{w}\mathbf{v}, \qquad \mathbf{w} \in \mathfrak{Im}\,\mathbb{H}.
\end{aligned}
\tag{23.26}
$$

The consequences of this calculation are pursued in the following key result in the subject.

Theorem 23.11. *Let* $\mathbf{q}$ *be a unit quaternion. Then* $R[\mathbf{q}]$ *represents a counterclockwise rotation by the angle* $2\arccos(\mathfrak{Re}\,\mathbf{q})$*, measured between* 0 *and* 2π*, with axis pointing in the direction of* $\mathfrak{Im}\,\mathbf{q}$*.*

The qualification *counterclockwise* is to be interpreted when one looks downwards from a "mast" pointing in the direction of $\mathfrak{Im}\,\mathbf{q}$ onto the perpendicular plane of rotation. This is equivalent to the "right-hand corkscrew rule."

Proof. Write $\mathbf{v} = \mathfrak{Im}\,\mathbf{q}$ and $\mathbf{q} = a + \mathbf{v} \in \mathscr{U}$. There is a unique number θ, with $0 \leqslant \theta \leqslant 2\pi$ such that $a = \cos(\theta/2)$, and so $|\mathbf{v}| = \sin(\theta/2) \geqslant 0$.

If $\mathbf{v} = 0$ then $\mathbf{q} = \pm 1$ and θ is 0 or 2π, which gives rise to the identity rotation, and accords with the statement of the theorem. Hence, we may assume that $|\mathbf{v}| \neq 0$. Observe too that equation (23.26) implies that

$$
R[\mathbf{q}]\mathbf{v} = a^2\mathbf{v} + |\mathbf{v}|^2\mathbf{v} = \mathbf{v}.
\tag{23.27}
$$

Next, choose a unit imaginary quaternion $\mathbf{w}_1$ perpendicular to $\mathbf{v}$ in $\mathfrak{Im}\,\mathbb{H} = \mathbb{R}^3$. Then $\mathbf{w}_1 \cdot \mathbf{v} = 0$ and $\mathbf{v}\mathbf{w}_1 = -\mathbf{w}_1\mathbf{v} = \mathbf{v} \times \mathbf{w}_1$. If we set

$$
\mathbf{w}_3 = \frac{1}{|\mathbf{v}|}\mathbf{v}, \qquad \mathbf{w}_2 = \mathbf{w}_3\mathbf{w}_1,
$$

then $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is an orthonormal basis of $\mathbb{R}^3$ with $\mathbf{w}_1 \times \mathbf{w}_2 = \mathbf{w}_3$. Moreover, using (23.26),

$$
\begin{aligned}
R[\mathbf{q}]\mathbf{w}_1 &= \mathbf{w}_1 \cos^2\frac{\theta}{2} + 2\mathbf{v}\mathbf{w}_1\cos\frac{\theta}{2} - \mathbf{w}_1\sin^2\frac{\theta}{2} \\
&= \mathbf{w}_1\cos\theta + \mathbf{w}_2\sin\theta.
\end{aligned}
\tag{23.28}
$$

Similarly,

$$
\begin{aligned}
R[\mathbf{q}]\mathbf{w}_2 &= \mathbf{w}_2 \cos^2\frac{\theta}{2} + 2\mathbf{v}\mathbf{w}_2\cos\frac{\theta}{2} - \mathbf{w}_2\sin^2\frac{\theta}{2} \\
&= \mathbf{w}_2\cos\theta - \mathbf{w}_1\sin\theta,
\end{aligned}
\tag{23.29}
$$

exactly as in (23.5). Equations (23.27), (23.28), (23.29) assert that the linear transformation $R[\mathbf{q}]$ has matrix (23.4) relative to the orthonormal basis $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$, and is the rotation stated in the theorem. ▮

Note that both $\mathbf{q}$ and $-\mathbf{q}$ determine the same transformation; this is related to the appearance of half-angles $\theta/2$. Computationally, it is now an easy matter to use quaternions to model rotations, such as the one that Figure 23.3 attempts to represent:

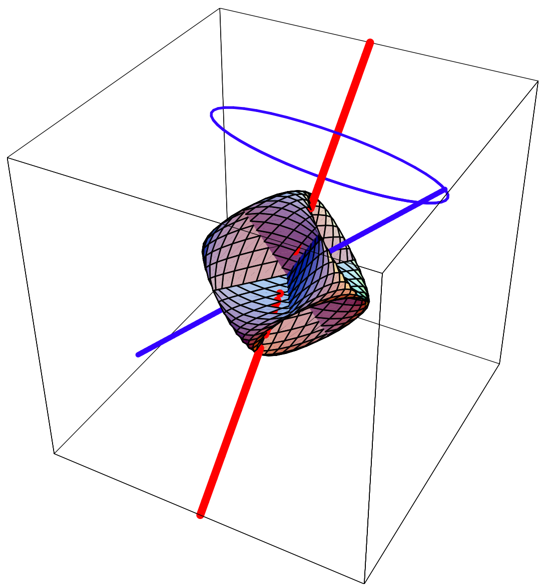

Figure 23.3: A rotation of the sine surface by $R[\frac{1}{\sqrt{2}} + \frac{1}{2}\mathbf{j} + \frac{1}{2}\mathbf{k}]$

The mapping

(23.30)
$$\begin{array}{ccc} \mathscr{U} & \longrightarrow & \mathbf{SO}(3) \\ \mathbf{q} & \mapsto & R[\mathbf{q}] \end{array}$$

satisfies the important property

(23.31)
$$R[\mathbf{q}_1\mathbf{q}_2] = R[\mathbf{q}_1] \circ R[\mathbf{q}_2].$$

This follows from its definition (23.24), since $R[\mathbf{q}_1\mathbf{q}_2]$ is the transformation

$$\mathbf{w} \mapsto (\mathbf{q}_1\mathbf{q}_2)\mathbf{w}\overline{(\mathbf{q}_1\mathbf{q}_2)} = \mathbf{q}_1(\mathbf{q}_2\mathbf{w}\overline{\mathbf{q}}_2)\overline{\mathbf{q}}_1.$$

Equation (23.31) characterizes what is called a ***group homomorphism***, a mapping from one group to another that preserves the respective multiplication rules. (For this topic, we recommend [Her, chapter 2].)

At this stage, we choose to write $R[\mathbf{q}]$ explicitly as a 3×3 orthogonal matrix. Referring to Lemma 23.2, this matrix is obtained by conjugating (23.4) by the orthogonal matrix whose columns are the orthonormal triple $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$. The result is well known and is computed in Notebook 23:

Proposition 23.12. *Let* $\mathbf{q} = (a, b, c, d)$ *be a unit quaternion. Then the rotation* $R[\mathbf{q}]$ *has matrix*

$$\begin{pmatrix} a^2+b^2-c^2-d^2 & 2bc-2ad & 2ac+2bd \\ 2bc+2ad & a^2-b^2+c^2-d^2 & 2cd-2ab \\ 2bd-2ac & 2ab+2cd & a^2-b^2-c^2+d^2 \end{pmatrix}. \tag{23.32}$$

In Notebook 23, we also verify directly that this matrix has determinant 1, even though this fact is a consequence of the proof of Theorem 23.11. It also follows from (23.26) and a continuity argument, as $\det R[\mathbf{q}]$ cannot jump from 1 to -1.

It is tempting to say that (23.32) is completely determined by the imaginary part $\mathfrak{Im}\,\mathbf{q} = (b, c, d)$ by defining

$$a = \sqrt{1-b^2-c^2-d^2}.$$

This is not quite true, because $\mathfrak{Re}\,\mathbf{q}$ may be negative, and changing the overall sign of $\mathbf{q}$ to adjust for this has the effect of reversing the sign of $\mathbf{v}$ and the angle of rotation. Nevertheless, there is a very natural way of converting a 3-vector into a rotation, and we describe this in the next section.

As an application of Theorem 23.11, consider the composition of two rotations $R[\mathbf{q}_1], R[\mathbf{q}_2]$ about axes $\mathbf{v}_1, \mathbf{v}_2$ with respective angles θ_1, θ_2. The combined rotation (23.31) is easiest to identify in the case in which the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal:

Corollary 23.13. *If* $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$*, the composition* $R[\mathbf{q}_1\mathbf{q}_2]$ *is a rotation through an angle* θ *with*

$$\cos\frac{\theta}{2} = \cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2},$$

and parallel to the axis

$$\sin\frac{\theta_1}{2}\cos\frac{\theta_2}{2}\,\mathbf{i} + \cos\frac{\theta_1}{2}\sin\frac{\theta_2}{2}\,\mathbf{j} + \sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2}\,\mathbf{k}.$$

Proof. Without loss of generality, we may assume that $\mathbf{v}_1, \mathbf{v}_2$ are parallel to the coordinate axes x, y. Then

$$\mathbf{q}_1 = a_1 + b_1\mathbf{i}, \qquad \mathbf{q}_2 = a_2 + b_2\mathbf{j},$$

where $a_i = \cos(\theta_i/2)$ and $b_i = \sin(\theta_i/2)$. Thus,

$$\mathbf{q}_1\mathbf{q}_2 = a_1a_2 \ + \ b_1a_2\,\mathbf{i} + a_1b_2\,\mathbf{j} + b_1b_2\,\mathbf{k},$$

and we can obtain the required formulas, bearing in mind that $\mathbf{q}_1\mathbf{q}_2$ is again a *unit* quaternion. ∎

23.4 Imaginary Quaternions and Rotations

The secret of using vectors to achieve rotations relies on the following fact, which (only) at first sight appears to have nothing to do with quaternions. Any rotation can be represented by a vector $\mathbf{x}$ in $\mathbb{R}^3$ with $|\mathbf{x}| \leqslant \pi$, by means of the following rule:

> *the axis of rotation is in the direction given by* $\mathbf{x}$*, and the angle of rotation is* $|\mathbf{x}|$ *radians counterclockwise.*

The representation above is almost unique, but there is an inevitable ambiguity, this time only if $|\mathbf{x}| = \pi$. (Recall that, in this chapter, $|\mathbf{x}|$ denotes the usual Euclidean norm $\|\mathbf{x}\|$.) For in this case, both $\mathbf{x}$ and $-\mathbf{x}$ represent the same rotation by π radians or 180°. This is because a rotation of θ counterclockwise about $\mathbf{x}$ is the same thing as a rotation of $-\theta$ or $2\pi - \theta$ counterclockwise about $-\mathbf{x}$, and the two coincide if $\theta = \pi$.

The description above incorporates a significant fact. To explain it, we first define the *closed ball*

$$B^3(\pi) = \{\mathbf{x} \in \mathbb{R}^3 \mid |\mathbf{x}| \leqslant \pi\}, \tag{23.33}$$

consisting of the sphere $S^2(\pi)$ of radius π, together with all its interior points. We are therefore saying that the set $\mathbf{SO}(3)$ of rotations can be thought of as the set $B^3(\pi)$ in which antipodal points on the boundary sphere $S^2(\pi)$ are identified. For if $|\mathbf{x}| = \pi$, then $\mathbf{x}$ and $-\mathbf{x}$ are *different* points in $B^3(\pi)$ that define the *same* rotation. The idea is that a bookworm living inside the ball that tunnels to the boundary does not get outside the sphere, but re-emerges at the point inside diametrically opposite the point it left.

It is a consequence of the description above that the set $\mathbf{SO}(3)$ of rotations forms a topological space *without* boundary. Mathematically, the sort of operation that has been performed on $B^3(\pi)$ to obtain $\mathbf{SO}(3)$ is a higher-dimensional analogue of the type of identifications carried out in Section 11.2. The argument after Figure 11.16 on page 346 shows that the disk $B^2(\pi)$ in $\mathbb{R}^2$ with points of its boundary circle $S^1(1)$ identified gives rise to the cross cap surface. The latter is a realization of the real projective plane $\mathbb{RP}^2$, which parametrizes straight lines passing through the origin in $\mathbb{R}^3$. The same argument can be repeated to see that the set $\mathbf{SO}(3)$ is topologically the real projective 3-space $\mathbb{RP}^3$ parametrizing lines through the origin in $\mathbb{R}^4$; see Exercise 12.

To understand the link with quaternions, we need to represent analytically the rotation defined by the vector $\mathbf{x}$. We can do this quickly using the machinery of the previous section. Let $\ell = |\mathbf{x}|$ denote the required angle of rotation, this time with $0 < \ell \leqslant \pi$. Set

$$a = \cos\frac{\ell}{2} \geqslant 0, \qquad \mathbf{v} = \left(\sin\frac{\ell}{2}\right)\frac{\mathbf{x}}{\ell}. \tag{23.34}$$

Then $\mathbf{q} = a+\mathbf{v}$ is a unit quaternion, whose imaginary part $\mathbf{v}$ has norm $\sin(\ell/2)$. Theorem 23.11 implies that $R[\mathbf{q}]$ is the required rotation. Recalling the notation (23.10), we can use this to establish

Corollary 23.14. *Let* $\mathbf{x} \in B^3(\pi)$. *The matrix of the rotation* $R[\mathbf{q}]$ (*with axis* $\mathbf{x}$ *and counterclockwise angle* $|\mathbf{x}|$) *equals* $\exp A[\mathbf{x}]$.

This result is effectively Theorem 23.4(ii), whose proof we now give.

Proof. Having fixed $\mathbf{x}$, we use (23.34), and temporarily abbreviate

$$A = A[\mathbf{v}] = \frac{\sin(\ell/2)}{\ell} A[\mathbf{x}].$$

Bearing in mind the form of equation (23.13), we shall express A^2 using quaternion multiplication. Assume first that $\mathbf{w}$ is perpendicular to $\mathbf{v}$. Then

$$\begin{aligned} 2A^2\mathbf{w} = 2A(\mathbf{v}\times\mathbf{w}) &= \mathbf{v}\times(2\mathbf{v}\times\mathbf{w}) \\ &= \mathbf{v}\times(\mathbf{v}\mathbf{w}-\mathbf{w}\mathbf{v}) \\ &= \frac{1}{2}\Big(\mathbf{v}(\mathbf{v}\mathbf{w}-\mathbf{w}\mathbf{v}) - (\mathbf{v}\mathbf{w}-\mathbf{w}\mathbf{v})\mathbf{v}\Big) \\ &= -\mathbf{v}\mathbf{w}\mathbf{v} - |\mathbf{v}|^2\mathbf{w}. \end{aligned} \tag{23.35}$$

Since $\mathbf{v}\mathbf{w}-\mathbf{w}\mathbf{v} = 2\mathbf{v}\mathbf{w}$ and $a^2+|\mathbf{v}|^2 = 1$, we may re-write (23.26) as

$$R[\mathbf{q}]\mathbf{w} = \mathbf{w} + 2aA\mathbf{w} + 2A^2\mathbf{w}.$$

This relation also holds when $\mathbf{w}$ is parallel to $\mathbf{v}$ because of (23.27). In conclusion,

$$\begin{aligned} R[\mathbf{q}] &= I_3 + 2\cos\frac{\ell}{2}A + 2A^2 \\ &= I_3 + \frac{\sin\ell}{\ell}A[\mathbf{x}] + \frac{1-\cos\ell}{\ell^2}A[\mathbf{x}]^2, \end{aligned}$$

as expected. ∎

It follows that

$$\mathbf{SO}(3) = \{\exp A[\mathbf{x}] \mid |\mathbf{x}| \leqslant \pi\},$$

so that the composition $\exp\circ A$ maps the closed ball $B^3(\pi)$ *onto* $\mathbf{SO}(3)$.

The descriptions we have seen make it easy to model rotations, a careful choice of which can be used to view the details of a surface. Figure 23.4 is a single frame of an animation from Notebook 23 of a truncated view of the sine surface, designed to reveal its inners (the axis of rotation is visible). Lines of self-intersections meet in a more complicated singularity at the origin.

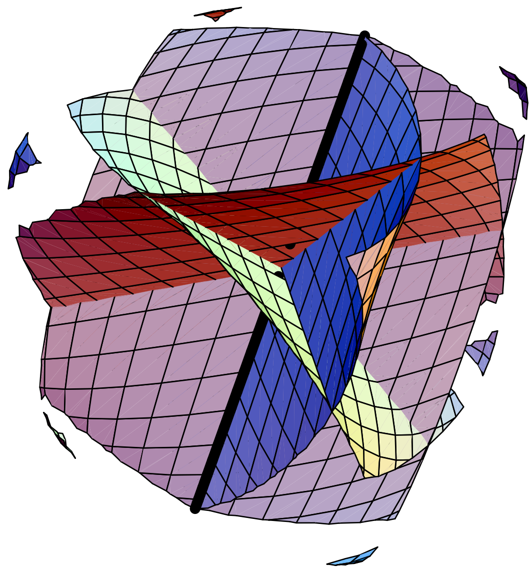

Figure 23.4: Viewing singularities of the sine surface

23.5 Rotation Curves

The theory of the preceding section taught us to associate to any vector $\mathbf{x}$ in the ball $B^3(\pi)$ a rotation in space, and to represent this rotation by the orthogonal matrix $\exp A[\mathbf{x}]$. In practice, this is best calculated using the formula of Theorem 23.4, which we now know arises from the unit quaternion

$$\mathbf{q} = \cos(\ell/2) + \frac{\sin(\ell/2)}{\ell}\mathbf{x}, \qquad \ell = |\mathbf{x}| \leqslant \pi, \tag{23.36}$$

defined by (23.34). The only ambiguity occurs when $|\mathbf{x}| = \pi$ for then $\pm\mathbf{x}$ give rise to the same unit quaternion $\mathbf{q}$.

Let $\mathbf{c}_j$ denote the j^{th} column of the matrix (23.32) representing (23.36). It is the image of the j^{th} basis element (from the list (23.12)) under the rotation. Since $a^2 + b^2 + c^2 + d^2 = 1$, we have

$$\mathbf{c}_1 = \begin{pmatrix} -2c^2 - 2d^2 + 1 \\ 2bc + 2ad \\ 2bd - 2ac \end{pmatrix} = R[\mathbf{q}] \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

The fact that $R[\mathbf{q}]$ is a rotation translates into the identity

$$\mathbf{c}_1 \times \mathbf{c}_2 = \mathbf{c}_3. \tag{23.37}$$

In particular, it means that $R[\mathbf{q}]$ is completely determined by the first two columns $\mathbf{c}_1, \mathbf{c}_2$.

We can now work backwards, and attempt to construct (23.32), $R[\mathbf{q}]$ and ultimately $\mathbf{x}$ itself from the columns of (23.32). If we start from the first column, the only restriction is that it be a *unit* vector, so

$$\mathbf{c}_1 \in S^2(1).$$

However, having fixed $\mathbf{c}_1$, the next column $\mathbf{c}_2$ must be chosen perpendicular to $\mathbf{c}_1$. This having been done, $\mathbf{c}_3$ is completely determined by (23.37).

Let $\mathcal{M} = S^2(1)$. If $\mathbf{u} \in \mathcal{M}$, then the orthogonal complement

$$\mathbf{u}^\perp = \{\mathbf{w} \in \mathbb{R}^3 \mid \mathbf{u} \cdot \mathbf{w} = 0\}$$

can be identified with the tangent space $\mathcal{M}_{\mathbf{u}}$. In general, the set

$$\{(\mathbf{u}, \mathbf{w}) \mid \mathbf{u} \in \mathcal{M},\ \mathbf{w} \in \mathcal{M}_{\mathbf{u}}\}$$

is called the ***tangent bundle*** of the surface $\mathcal{M}$, and is discussed in a more general setting in the next chapter (within Section 24.5). But for the moment, it leads to yet another description of the set $\mathbf{SO}(3)$.

Proposition 23.15. *The set* $\mathbf{SO}(3)$ *is in bijective correspondence with the* ***unit tangent bundle***

$$\{(\mathbf{u}, \mathbf{w}) \mid \mathbf{u}, \mathbf{w} \in S^2(1),\ \mathbf{u} \cdot \mathbf{w} = 0\}$$

of the unit sphere.

Note the symmetry between $\mathbf{u}$ and $\mathbf{w}$ in this definition.

Proposition 23.15 is relevant to the next topic. At the start of this book, on page 5, we defined a curve in $\mathbb{R}^n$, and in Chapters 7 we specialized to the case $n = 3$. We are now in a position to adapt this to the case in which $\mathbb{R}^3$ is replaced by $\mathbf{SO}(3)$.

Definition 23.16. *A* ***rotation curve*** *is a mapping* $\boldsymbol{\gamma}\colon [a, b] \to \mathbf{SO}(3)$ *of the form*

$$\boldsymbol{\gamma}(t) = \exp A[\boldsymbol{\alpha}(t)],$$

where $\boldsymbol{\alpha}\colon [a, b] \to \mathbb{R}^3$ *is a parametrized curve whose image lies in* $B^3(\pi)$. *The rotation curve is* ***closed*** *if* $\boldsymbol{\gamma}(a) = \boldsymbol{\gamma}(b)$.

We begin with two basic examples of such curves.

(i) Fix a unit vector $\mathbf{u}$, and set

$$\boldsymbol{\alpha}(t) = t\,\mathbf{u}, \qquad -\pi \leqslant t \leqslant \pi.$$

Then $\boldsymbol{\gamma}(t) = \exp A[t\mathbf{u}]$ is a rotation, by an angle t, about a fixed axis parallel to $\mathbf{u}$. The curve $\boldsymbol{\alpha}$ is a diameter of $B^3(\pi)$, but $\boldsymbol{\gamma}$ is closed since $\exp A[-\pi\mathbf{u}] = \exp A[\pi\mathbf{u}]$.

(ii) Fix an orthonormal pair $\mathbf{u}_1, \mathbf{u}_2$ of vectors, and set

$$\boldsymbol{\beta}(t) = \Big(\mathbf{u}_1 \cos\frac{t}{2} + \mathbf{u}_2 \sin\frac{t}{2}\Big)\pi, \qquad -\pi \leqslant t \leqslant \pi.$$

The trace of $\boldsymbol{\beta}$ is half of a great circle lying on the boundary $S^2(\pi)$ of $B^3(\pi)$, and *every* element $\boldsymbol{\delta}(t) = \exp A[\boldsymbol{\beta}(t)]$ is a rotation of 180°. The rotation curve $\boldsymbol{\delta}$ is again closed since $\boldsymbol{\beta}(\pi) = \pi\mathbf{u}_2 = -\boldsymbol{\beta}(-\pi)$.

These two examples are almost identical, though this fact is partially hidden by our preference for imaginary quaternions; matters become clearer when we revert to unit quaternions. In case (i), the associated unit quaternion is

$$\mathbf{q}(t) = \cos\frac{t}{2} + \mathbf{u}\sin\frac{t}{2}, \tag{23.38}$$

whereas for (ii), $\boldsymbol{\beta}(t)$ is itself a unit imaginary quaternion. Indeed, $\boldsymbol{\beta}(t)$ has the same form as (23.38) with $\mathbf{u}_1, \mathbf{u}_2$ in place of $1, \mathbf{u}$; if we choose $\mathbf{u} = \overline{\mathbf{u}}_1\mathbf{u}_2$, then

$$\boldsymbol{\beta}(t) = \mathbf{u}_1\mathbf{q}(\theta).$$

This means that

$$\boldsymbol{\delta}(t) = R[\mathbf{u}_1]\,\boldsymbol{\gamma}(t),$$

and the two rotation curves differ by the fixed 180° rotation $R[\mathbf{u}_1]$. In technical language, *left translation* within the group $\mathscr{U}$ converts a diameter of $B^3(\pi)$ into half a great circle.

Surprisingly complicated examples can be obtained by starting from a curve whose trace lies in the intersection of $B^3(\pi)$ with a plane through the origin. Using standard coordinates in $\mathbb{R}^3$, consider the ellipse

$$\boldsymbol{\alpha}[a,b](t) = \big(0,\ a\cos t,\ b\sin t\big)\pi, \qquad -\pi \leqslant t \leqslant \pi, \tag{23.39}$$

where a, b are fixed positive numbers such that $a^2\cos^2 t + b^2\sin^2 t \leqslant 1$ for all t. Let $\boldsymbol{\gamma}(t) = \exp A[\boldsymbol{\alpha}[a,b](t)]$; both $\boldsymbol{\alpha}[a,b]$ and $\boldsymbol{\gamma}$ are closed curves. To describe $\boldsymbol{\gamma}$, it suffices to specify the first two columns $\boldsymbol{\gamma}(t)\mathbf{e}_1$ and $\boldsymbol{\gamma}(t)\mathbf{e}_2$ in accordance with the discussion leading to Proposition 23.15. As t varies, these generate two curves lying on $S^2(\pi)$.

This example is animated in Notebook 23 and illustrated in Figure 23.5. One can imagine a rigid body with two perpendicular arms that are constrained to follow the respective spherical curves. In the frame shown, the rotation is about to pass through a "vertex" in which one column achieves an extreme value, and the second passes through the chicane in the figure eight.

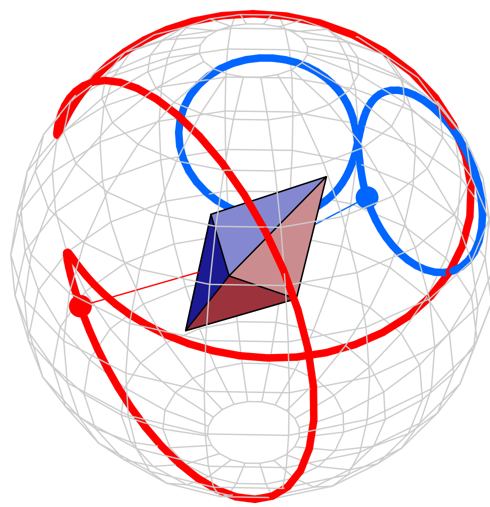

Figure 23.5: A rotation curve specified by its columns

Figure 23.6 concerns the case of (23.39) in which $a = b$ and

$$\boldsymbol{\gamma}[a](t) = \exp(A[\boldsymbol{\alpha}[a,a]]), \qquad 0 \leqslant t \leqslant 2\pi.$$

From example (ii) above, $\boldsymbol{\gamma}[1]$ represents a rotation curve corresponding to revolving two turns (4π radians or 720^{o}) around a fixed axis. At the other extreme, $\boldsymbol{\gamma}[0]$ is the "identity curve," whose trace consists of the single point $I_3 \in \mathbf{SO}(3)$. Figure 23.6 displays the intermediate curves $\boldsymbol{\gamma}[1-a]$ with $a = n/17$ and $1 \leqslant n \leqslant 16$. In the first frame, the straightforward 720^{o} rotation has been deformed slightly to $\boldsymbol{\gamma}[16/17]$ so that the axis begins to "wobble." By the time we get down to $\boldsymbol{\gamma}[4/17]$, represented by the first frame in the last row, a rigid body subject to this family of rotations will merely rock slowly about an axis lullaby-fashion (imagine Figure 23.5 with the more tranquil background).

The family

$$a \mapsto \boldsymbol{\gamma}[1-a], \qquad 0 \leqslant a \leqslant 1$$

is a continuous deformation from the "two turns curve" $\boldsymbol{\gamma}[1]$ to the identity $\boldsymbol{\gamma}[0]$. In the language of Section 6.3, it is a *homotopy* between $\boldsymbol{\gamma}[1]$ and $\boldsymbol{\gamma}[0]$. There are other ways of demonstrating the existence of such a homotopy, such as the so-called Dirac scissors trick (see [Pen, §11.3]). There exists no such homotopy if we begin with the closed rotation curve $\boldsymbol{\gamma}[1]$ defined for $0 \leqslant t \leqslant \pi$, which corresponds to a single revolution. This fact is expressed by saying that the topological space $\mathbf{SO}(3)$, which we have mentioned is equivalent to $\mathbb{RP}^3$, is not ***simply-connected*** — there exist closed curves that cannot be deformed to a point. It is however the case that the "double" of any closed curve can always be deformed to a point. These facts are explained in introductory texts to algebraic topology, such as [Arm].

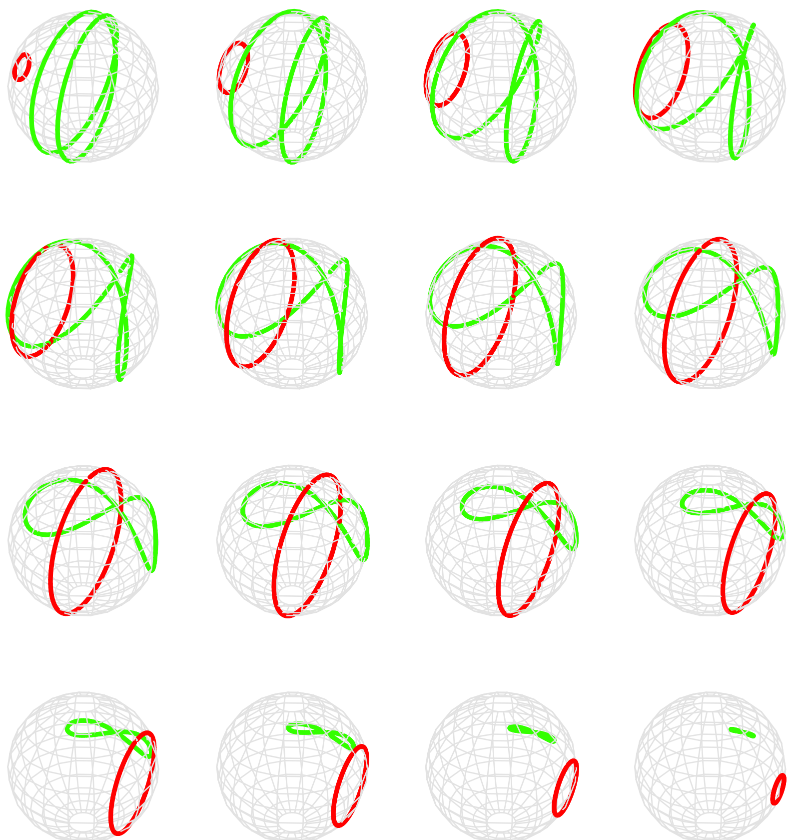

Figure 23.6: A homotopy of rotation curves $\boldsymbol{\gamma}[\frac{n}{17}]$

23.6 Euler Angles

The traditional way to specify a rotation is to break it up into the composition of three rotations about the three Cartesian axes. In practice, this is the method used in fairground machinery, or in a gyroscope to align a navigation system.

In order to provide a vivid explanation of Euler angles, we define them in relation to an aircraft taking off from runway 27L at Heathrow airport in London, on its way to Paris. Imagine first a Cartesian frame of reference $\mathbf{F}_0$ fixed in the center of the runway, with the x-axis due east, y-axis due north (which just happens to be perpendicular to the runway), and z (for "zenith") vertically upwards. As the aircraft passes by on its take-off run, its own frame of reference $\mathbf{F}_1$ (with x pointing to the nose and y to the port or left wing) is obtained from $\mathbf{F}_0$ by means of a rotation of 180° about the z-axis, to give the correct heading of 270°. After takeoff, the aircraft climbs at a constant angle or elevation of 20°; mathematically this is achieved by rotating $\mathbf{F}_1$ by $-20°$ about its y-axis, and produces a new frame of reference $\mathbf{F}_2$. At a height of 2000 feet, the aircraft banks left, by rolling $-30°$ about its nose-tail or x-axis, temporarily achieving the frame of reference $\mathbf{F}_3$.

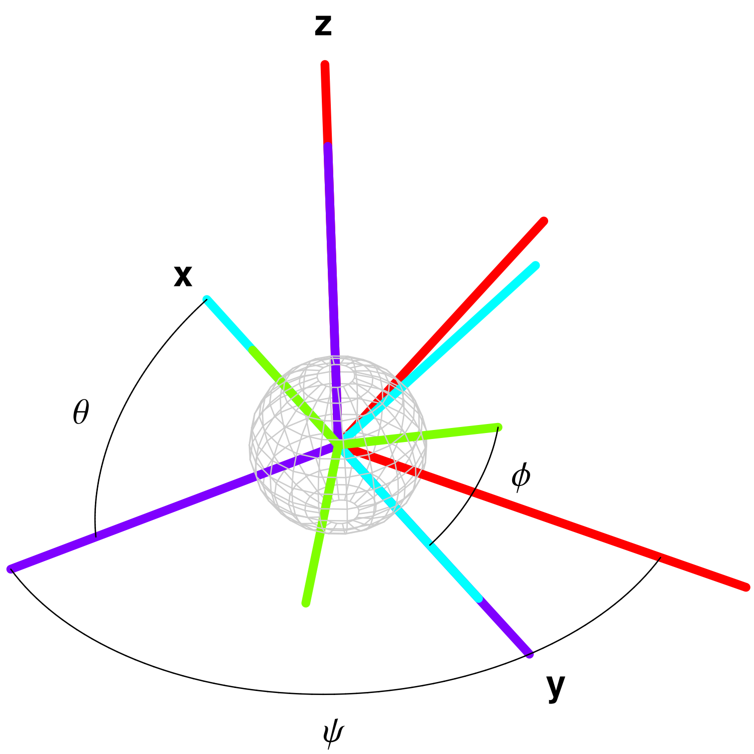

Figure 23.7: A sequence of rotations about axes $z \rightsquigarrow y \rightsquigarrow x$

To sum up, $\mathbf{F}_3$ has been obtained from $\mathbf{F}_0$ by means of a sequence of rotations through

$$\psi = 180^{\circ}, \quad \theta = -20^{\circ}, \quad \phi = -30^{\circ} \tag{23.40}$$

about the z-, y-, x-axes respectively (our notation is similar to that of [Kuipers, 4.4]). The triple $(\psi, \theta\, \phi)$ comprises the ***Euler angles*** of the single rotation P needed to move $\mathbf{F}_0$ to $\mathbf{F}_3$, and can be regarded as yet another way of representing a rotation by a 3-vector (albeit one whose three components are angles).

To be more precise, we can interpret each frame $\mathbf{F}_i$ as the orthonormal triple of vectors defining the frame's axes, and translate all the frames to a common origin. In this way, $\mathbf{F}_3$ is obtained from $\mathbf{F}_0$ by the composition

$$R[a_3 + b_3\mathbf{i}] \circ R[a_2 + b_2\mathbf{j}] \circ R[a_1 + b_1\mathbf{k}]$$

of rotations, where

$$a_3 = \cos\frac{\phi}{2}, \quad a_2 = \cos\frac{\theta}{2}, \quad a_1 = \cos\frac{\psi}{2}.$$

One can easily compute P using Corollary 23.13.

The definition of the Euler angles depends on the sequence of axes chosen to perform the respective three rotations. The above choice $z \rightsquigarrow y \rightsquigarrow x$ (meaning first z then y then x) is illustrated in Figure 23.7 but with a different choice of angles ($\psi = -2\pi/3$, $\theta = -\pi/3$, $\phi = +\pi/4$ in radians). At each step the new axes are shortened to help identify them, but the figure is best seen on screen in color. One can attempt to use any other sequence to obtain a given rotation, such as $x \rightsquigarrow y \rightsquigarrow x$, though it is useless if the same axis appears next to itself in the sequence. It follows that there is a total of $3 \cdot 2 \cdot 2 = 12$ choices for the triple, but essentially only two types: those in which the three axes are different and those in which the first and third coincide. As it happens, the standard rotation package within *Mathematica* uses the sequence $z \rightsquigarrow x \rightsquigarrow z$ (see Notebook 23).

Euler's theorem states that any element in $\mathbf{SO}(3)$ can be decomposed as the composition of three rotations about mutually orthogonal axes, in either of the above ways we care to choose. Using the theory from Section 23.3, it amounts to the following factorization result for unit quaternions. First, recall the notation (23.20), and consider the circle

$$\mathscr{U}_{\mathbf{v}} = \{a + b\mathbf{v} \mid a^2 + b^2 = 1\} \tag{23.41}$$

formed by intersecting the subspace $\langle\, 1,\ \mathbf{v}\,\rangle$ with $\mathscr{U}$.

Theorem 23.17. *Let $\mathbf{q} \in \mathscr{U}$. Then there exist*

(i) $\mathbf{q}_1 \in \mathscr{U}_{\mathbf{i}}$, $\mathbf{q}_2 \in \mathscr{U}_{\mathbf{j}}$, $\mathbf{q}_3 \in \mathscr{U}_{\mathbf{k}}$ *such that* $\mathbf{q} = \mathbf{q}_3\mathbf{q}_2\mathbf{q}_1$;

(ii) $\mathbf{p}_1 \in \mathscr{U}_{\mathbf{i}}$, $\mathbf{p}_2 \in \mathscr{U}_{\mathbf{j}}$, $\mathbf{p}_3 \in \mathscr{U}_{\mathbf{i}}$ *such that* $\mathbf{q} = \mathbf{p}_3\mathbf{p}_2\mathbf{p}_1$.

We omit the proof, referring the reader to [Kuipers, chapter 8]. A first step consists in the characterization of those quaternions that can be represented in the form $\mathbf{q}_1\mathbf{q}_2$ where $\mathbf{q}_1 \in \mathscr{U}_{\mathbf{i}}$ and $\mathbf{q}_2 \in \mathscr{U}_{\mathbf{j}}$ (see Exercise 8). To explain the nature of the problem, we prove instead the following simpler result.

Lemma 23.18. *Given $\mathbf{v} \in \mathfrak{Im}\,\mathbb{H}$, there exist $a, b, c \in \mathbb{R}$ such that*

$$\mathfrak{Im}[(a + \mathbf{i})(b + \mathbf{j})(c + \mathbf{k})] = \mathbf{v}.$$

Proof. Writing $\mathbf{v} = r\mathbf{i} + s\mathbf{j} + t\mathbf{k}$ and expanding, we need to solve the system

$$\left\{\begin{array}{rcl} a + bc &=& r, \\ -b + ac &=& s, \\ c + ab &=& t, \end{array}\right. \tag{23.42}$$

that does not quite have cyclic symmetry. Eliminating b, we obtain

$$\left\{\begin{array}{rcl} a(1+c^2) &=& cs + r, \\ c(1+a^2) &=& as + t. \end{array}\right.$$

Eliminating c gives the quintic equation

$$a^5 - r\,a^4 + 2a^3 + (st - 2r)a^2 + (t^2 - s^2 + 1)a - r - st = 0,$$

which has at least one real root a. The last two equations in (23.42) are then linear and can be solved for b and c. ∎

There are well-documented disadvantages of the use of Euler angles to describe rotations. Not only do they depend very much on the choice of axis sequence, but are subject to an analog of ***gimbal lock*** of an inertial guidance system. Fear of gimbal lock was the bane of mission control during the Apollo moon landing and other space flights. One manifestation of the difficulty is that when an Euler angle approaches $\pi/2$, its cosine approaches 0, and it is impossible for a computer to perform division by $\cos\theta$. Mechanically, one can partly understand the problem by supposing that our aircraft is able (quickly after takeoff) to climb vertically, so that θ has become -90° in (23.40). In these circumstances, the aircraft's heading has been "lost," the third rotation could have been achieved by changing the direction of the runway, and ψ, ϕ are no longer independent parameters.

The advantage of our earlier description of rotations based on imaginary quaternions is that the mapping $B^3(\pi) \to \mathbf{SO}(3)$ defined by

$$\boldsymbol{\rho}(\mathbf{x}) = \exp A[\mathbf{x}] = R[\mathbf{q}]$$

is *regular* in the sense that the partial derivatives $\boldsymbol{\rho}_x, \boldsymbol{\rho}_y, \boldsymbol{\rho}_z$ are always linearly independent (as 3×3 matrices, or elements of $\mathbb{R}^9$), in the spirit of Definition 19.16 on page 606. This is *not* true of the mapping $\boldsymbol{\sigma}\colon \mathcal{R} \to \mathbf{SO}(3)$ defined by

$$\boldsymbol{\sigma}(\psi, \theta, \phi) = \mathbf{F}_3,$$

where $\mathcal{R} = [0, 2\pi] \times [-\pi/2, \pi/2] \times [-\pi, \pi]$ is the usual domain of definition for the Euler angles. Indeed, we verify in Notebook 23 that the set

$$\left\{ \boldsymbol{\sigma}\Big(\psi, \frac{\pi}{2}, \phi\Big) \;\Big|\; \psi, \phi \in \mathbb{R} \right\}$$

is a curve and not a surface in $\mathbf{SO}(3)$.

23.7 Further Topics

We conclude this chapter with some more applications. The first two link to the themes of Chapters 7 and 15 respectively.

Frenet Frames Revisited

We can associate to each point of a space curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ a rotation matrix $\mathbf{F}(t)$ as follows. The parameter t need not be arc length. The columns of $\mathbf{F}(t)$ are the unit tangent, normal and binormal vectors $\mathbf{T}(t), \mathbf{N}(t), \mathbf{B}(t)$ at the point $\boldsymbol{\alpha}(t)$. These are mutually orthogonal and by definition,

$$\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t).$$

It follows that $\mathbf{F}(t)$ lies in $\mathbf{SO}(3)$, and represents the rotation necessary to transform the standard frame of reference $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to $\mathbf{F}(t)$, in analogy to the aeronautical example of the previous section.

The assignment of the unit tangent vector $\mathbf{T}(t)$, thought of as a point on $S^2(1)$, to $\boldsymbol{\alpha}(t)$ is a type of Gauss map for the curve, although it could be argued that using $\mathbf{N}(t)$ in place of $\mathbf{T}(t)$ is closer to the Gauss definition for surfaces. The fact is that we may as well associate the *whole Frenet frame* $\mathbf{F}(t)$ to a point $\boldsymbol{\alpha}(t)$ where $\boldsymbol{\kappa}(t) \neq 0$. The resulting assignment

$$t \mapsto \boldsymbol{\alpha}(t) \mapsto \mathbf{F}(t), \tag{23.43}$$

ultimately mapping the trace of $\boldsymbol{\alpha}$ to $\mathbf{SO}(3)$, is an example of a "higher order" Gauss map, of the sort that is much studied in current research.

The composition (23.43) describes a rotation curve. The latter may be closed even if $\boldsymbol{\alpha}$ is not, and the best illustration of this phenomenon is the helix

$$\mathsf{helix}[a,b,c](t) = \big(a\cos t,\ b\sin t,\ ct\big)$$

(see Section 7.3), which is circular if $a = b$. Since

$$\mathsf{helix}[a,b,c](t+2\pi) = \mathsf{helix}[a,b,c](t) + (0,\ 0,\ 2\pi c),$$

it follows that $\mathbf{F}(t+2\pi) = \mathbf{F}(0)$, and

$$\mathbf{F}\colon [0, 2\pi] \longrightarrow \mathbf{SO}(3)$$

is a closed rotation curve.

A more subtle example is the cubic curve

$$\mathsf{twicubic}(t) = (t,\ t^2,\ t^3) \tag{23.44}$$

introduced on page 202. When $|t|$ is very large, we may ignore t and t^2 in comparison to t^3, and the curve approximates the vertical straight line $t \mapsto (0,0,t^3)$

for which $\mathbf{T}(t) = (0,0,1)$ independently of t (be it positive or negative). On the other hand, the curvature $\boldsymbol{\kappa}$[twicubic] computed on page 205 never vanishes. Thus, it is always possible to define $\mathbf{N}$ and $\mathbf{B}$, and one verifies that

$$\begin{aligned} \lim_{t\to\pm\infty} \mathbf{N}(t) &= (0,\, -1,\, 0), \\ \lim_{t\to\pm\infty} \mathbf{B}(t) &= (1,\, 0,\, 0) \end{aligned} \tag{23.45}$$

(see Exercise 10). As a consequence, adjoining the single point

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

makes $\mathbf{F}(t)$ into a smooth closed curve in $\mathbf{SO}(3)$. What is really going on here is that (23.44) can be extended to a mapping $\mathbb{RP}^1 \to \mathbb{RP}^3$ in the context of projective geometry (see Exercise 12), and $\mathbf{F}$ too has values in $\mathbb{RP}^3$.

Figure 23.8 shows the situation for Viviani's curve, which is somewhat of a special case, as the original curve already lies on the sphere. It displays the traces of $\mathbf{T}$ and $\mathbf{B}$ (the latter with cusps), in the style of previous figures. Despite first appearances, the unit tangent vector

$$\mathbf{T}(t) = \frac{\sqrt{2}}{\sqrt{3+\cos t}}\Big(\sin t,\, \cos t,\, \cos\frac{t}{2}\Big) \tag{23.46}$$

is not itself the intersection of the sphere with a circular cylinder (see Exercise 9).

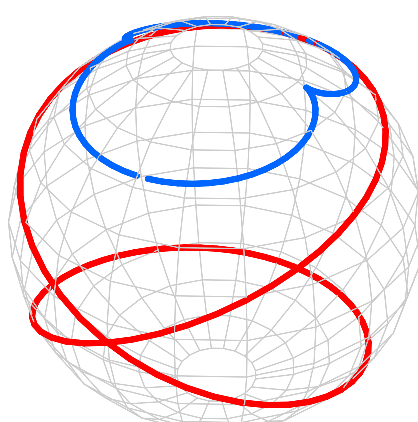

Figure 23.8: The moving Frenet frame on Viviani's curve

It is tempting to think that *any* curve $\boldsymbol{\gamma}\colon (a,b) \to \mathbf{SO}(3)$ arises as the Frenet frame $\mathbf{F}(t)$ of a suitable space curve $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^3$. The Frenet formulas, in the version of Theorem 7.13 on page 203, show that this is not however the case, since $\mathbf{T}'(t)$ must be parallel to the second column $\mathbf{N}(t)$ of $\boldsymbol{\gamma}(t)$. Thus, the rotation curve

$$\mathbf{F} : (a,b) \to \mathbf{SO}(3)$$

defined by the Frenet frame of $\boldsymbol{\beta}$ is completely determined by the unit tangent mapping $\mathbf{T}\colon (a,b) \to S^2(1)$. This provides a significant constraint on the trace of $\mathbf{F}$, which is best understood using the description of $\mathbf{SO}(3)$ given by Proposition 23.15.

Generalized Surfaces of Revolution

On page 471 we mentioned Darboux's point of view, whereby a surface of revolution is viewed as the surface swept out by the one-parameter group of motions corresponding to rotation about a fixed axis. This motivated the definition of generalized helicoid, for which a similar interpretation holds. Now that we have developed machinery to describe a general rotation curve

$$\boldsymbol{\gamma}\colon (a,b) \to \mathbf{SO}(3),$$

we could equally well apply this to a given space curve $\boldsymbol{\beta}\colon (c,d) \to \mathbb{R}^3$, so as to obtain the chart

$$\mathbf{x}(u,v) = \boldsymbol{\gamma}(u)\boldsymbol{\beta}(v).$$

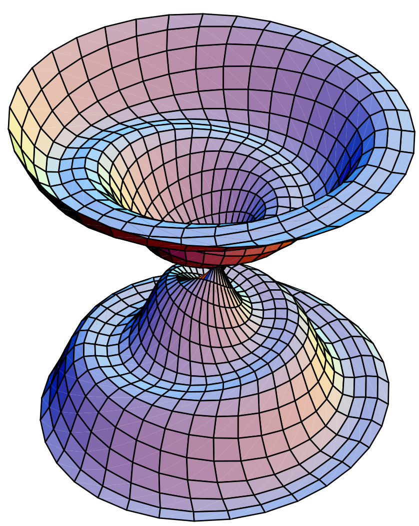

Figure 23.9: A surface generated by applying a rotation curve to a helix

A natural question is what happens when $\boldsymbol{\gamma}$ is one of the ellipses (23.39). Figure 23.1 was obtained by taking

$$\boldsymbol{\gamma}(u) = \exp A[(0,\ \tfrac{9}{10}\cos u,\ \tfrac{27}{100}\sin u)]$$

in the notation of (23.10), and $\boldsymbol{\beta}$ to be the straight line

$$\boldsymbol{\beta}(v) = (v,\ 0,\ 0).$$

Figure 23.9 applies a similar rotation curve to the space curve

$$\boldsymbol{\beta}(v) = \big(\cos 5v,\ \sin 5v,\ 10v\big).$$

Rotations in $\mathbb{R}^4$

The astute reader may question the extent to which quaternions are essential for an analysis of the rotation group. While it is true that Theorem 23.4 provided the main computational power of the chapter, one should not underestimate the mathematical significance of Theorem 23.11, particularly in extending the theory to $\mathbb{R}^n$. To finish, we merely state a generalization of Lemma 23.10.

Lemma 23.19. *Let $\mathbf{q}_1, \mathbf{q}_2 \in \mathscr{U}$. If $\mathbf{p} \in \mathbb{H}$ then $\mathbf{q}_1\mathbf{p}\overline{\mathbf{q}}_2$ is a quaternion with the same norm as $\mathbf{p}$. The resulting mapping*

$$\mathbf{p} \mapsto \mathbf{q}_1\mathbf{p}\overline{\mathbf{q}}_2 \tag{23.47}$$

is an orthogonal transformation of $\mathbb{R}^4$ and determines a matrix in $\mathbf{SO}(4)$.

Proof. It suffices to modify the step (23.25) for the previous proof to work. ▮

This leads to a description of $\mathbf{SO}(4)$ based on $\mathscr{U} \times \mathscr{U}$. Further to (23.30), there is an associated homomorphism of groups

$$\mathscr{U} \times \mathscr{U} \longrightarrow \mathbf{SO}(4).$$

Exactly two elements of $\mathscr{U} \times \mathscr{U}$ (namely, $(\mathbf{q}_1, \mathbf{q}_2)$ and $(-\mathbf{q}_1, -\mathbf{q}_2)$) map to the same rotation of $\mathbb{R}^4$. Elements of the type $(\mathbf{q}, \mathbf{q})$ really do map to $\mathbf{SO}(3)$, in the sense that they leave fixed the direction generated by $1 \in \mathbb{H}$, in the same way that (23.4) is effectively a rotation of $\mathbb{R}^2$ rather than $\mathbb{R}^3$.

The ability to describe a rotation by two separate elements $\mathbf{q}_1, \mathbf{q}_2$ is a special feature of the group of rotations in *four* dimensions. This led in the 1970s to a fuller understanding by mathematicians of gauge theories in theoretical physics, a resurgence of interest in quaternions [Atiyah], and unforeseen developments in pure mathematics.

23.8 Exercises

1. Prove that the matrix $A = A[\mathbf{x}]$ satisfies

(a) $A^3 = -\|\mathbf{x}\|^2 A$.

(b) $A^2 = \mathbf{x}\mathbf{x}^T - \|\mathbf{x}\|^2 I_3$.

2. Let $\mathbf{p}, \mathbf{q} \in \mathbb{H}$. Under what conditions on $\mathbf{p}$ and $\mathbf{q}$ is it true that $\mathbf{pq} = \mathbf{qp}$?

3. Show that the eight elements

$$1, \quad -1, \quad i, \quad j, \quad k, \quad -i, \quad -j, \quad -k$$

form a group under quaternion multiplication.

4. Let $P \in \mathbf{SO}(3)$, $P^2 \neq I_3$. Show that P represents a rotation about an axis parallel to the vector $\mathbf{x}$, where $A[\mathbf{x}] = P - P^T$.

5. Prove that if X, Y are skew-symmetric $n \times n$ matrices then so is $XY - YX$. Now suppose that $n = 3$ and that $X = A[\mathbf{x}]$, $Y = A[\mathbf{y}]$. Use (23.11) to verify that

$$XY - YX = A[\mathbf{x} \times \mathbf{y}],$$

and relate this fact to the Jacobi identity (24.46) on page 832.

6. Let $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ be a complex 2×2 matrix satisfying $A^T \overline{A} = I_2$, where $\overline{A}$ is the matrix obtained from A by complex-conjugating its four elements and I_2 is the identity. Show that the complex number $\det A = \alpha\delta - \beta\gamma$ has modulus one, and that if $\det A = 1$ then $\gamma = -\overline{\beta}$ and $\delta = \overline{\alpha}$.

7. Prove that the set of matrices

$$\mathbf{SU}(2) = \left\{ \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} \;\middle|\; |\alpha|^2 + |\beta|^2 = 1 \right\}$$

arising from the previous question is a group under matrix multiplication. By writing each complex entry in real and imaginary parts, show that its elements are in bijective correspondence with those of the group $\mathscr{U}$ of unit quaternions.

8. Show that a unit quaternion

$$\mathbf{q} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$$

can be written as a product $\mathbf{q} = \mathbf{q}_1\mathbf{q}_2$ with $\mathbf{q}_1 \in \mathscr{U}_{\mathbf{i}}$ and $\mathbf{q}_2 \in \mathscr{U}_{\mathbf{j}}$ if and only if $ad - bc = 0$. See (23.41) for the notation.

M **9.** Plot the projection of the curve (23.46) on the xy-plane, and verify that it is not a circle. Find analogous expressions for the normal and binormal vectors to Viviani's curve.

M **10.** Verify the limits (23.45), and plot the rotation curve $\mathbf{F}(t)$ for a twisted cubic.

M **11.** Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^3$ be a curve with speed $v = v(t)$ whose Frenet frame defines $\boldsymbol{\gamma}\colon (a,b) \to \mathbf{SO}(3)$. Use Theorem 7.13 to show that

$$\boldsymbol{\gamma}'(t) = \boldsymbol{\gamma}(t)\, A[\mathbf{x}(t)],$$

where $\mathbf{x}(t) = (-\boldsymbol{\tau}(t),\, 0,\, \boldsymbol{\kappa}(t))v$.

12. Define an equivalence relation $\sim$ on $S^n(1) \subset \mathbb{R}^{n+1}$ by writing $\mathbf{a} \sim \mathbf{b}$ if and only if $\mathbf{a} = \pm\mathbf{b}$. The resulting quotient space is called ***real projective space*** of dimension n and denoted by $\mathbb{RP}^n$. The case $n = 2$ was the subject of Section 11.5. Define $p\colon S^n(1) \mapsto \mathbb{RP}^n$ by $p(\mathbf{a}) = \{\mathbf{a}, -\mathbf{a}\}$. Deduce from Theorem 23.11 that there is a bijective correspondence between $\mathbf{SO}(3)$ and $\mathbb{RP}^3$.