AN INTRODUCTION TO QUATERNIONS WITH APPLICATION TO ROTATIONS

LARRY SUSANKA

Contents

Date: October 11, 2017.

1. Algebras

A (real) **algebra**[1] is a vector space V whose scalars are the real numbers $\mathbb{R}$ together with an additional operation on pairs of vectors, usually indicated as multiplication, with a few properties.

First, the product vw of the vectors v and w from V must produce a vector in V. This property is called closure, and applies to subsets too. If $S \subset V$ we say the product is closed in S if vw is in S whenever both v and w are in S.

Unless the vector space is the trivial one dimensional space, $\mathbb{R}$ itself, dot product on $\mathbb{R}^n$ produces a number, not a vector, so the familiar dot product is *not* an example of what we mean here.

The usual cross product on vectors in $\mathbb{R}^3$ *is* an example. The usual matrix multiplication on the vector space of n-by-n matrices is another.

Second, the two distributive laws must hold for this operation:

$$(v + w)z = vw + vz \quad \text{and} \quad z(v + w) = zv + zw$$

for any vectors v, w and z.

We adopt the usual practice of giving multiplication priority over addition, so $zv + zw = (zv) + (zw)$ and is not any of the other possible combinations, such as $(z(v + z))w$.

Both cross product and matrix multiplication satisfy these distributive laws.

Finally, if r and s are scalars (i.e. real numbers) and v and w are vectors then

$$(rs)(vw) = ((rs)v)w = (rv)(sw) = v((rs)w).$$

Cross product and matrix multiplication satisfy this one too.

Taking a look at the equation above we see three different multiplications, all indicated by "juxtaposition," i.e. putting the symbols to be multiplied next to each other. rs is the product of two real numbers. vw is the product of two vectors. rv is scalar multiplication of real number r by vector v. Believe it or not, in applications this rarely causes confusion.

An algebra is called associative if, for any vectors v, w and z

$$(vw)z = v(wz).$$

This property is important because without it the product of two or more terms, such as $vwzq$, is ambiguous. One must then include a lot of parentheses to dictate order. Matrix multiplication is associative, but cross product is not.

An algebra is called commutative if $vw = wv$ for any vectors v, w. Matrix multiplication of n-by-n matrices is not commutative, and neither is cross product. However the last operation satisfies $v \times w = -w \times v$, a property sometimes called "anticommutativity."

---

[1]Those who wish to minimize abstraction but still learn the basic facts can stick to the sections marked with (*) and skip the rest without too much much loss of continuity. Readers without much background in linear algebra, on the other hand, will likely find mention of vector spaces, linear transformation, basis, dimension, homomorphism and isomorphism to be puzzling, but can still understand many of the ideas in these notes by simply ignoring any sentence containing the words.

An algebra is called **unital** if there is a vector, sometimes denoted e, for which

$$ev = ve = v \quad \text{for all vectors } v.$$

This is not to be confused with the scalar, number 1. The multiplicative identity e is a vector, not a scalar, and is unique if it exists.

Matrix multiplication is unital, but cross product is not.

An algebra is called a **division algebra** if for every nonzero vector v and any vector w there is exactly one vector z and exactly one vector y for which $w = zv$ and $w = vy$.

For an associative unital division algebra the above condition is equivalent to the following condition: for every vector v there is a vector, usually denoted v^{-1}, for which $vv^{-1} = v^{-1}v = e$.

An algebra homomorphism from algebra V to algebra W is a linear transformation T from V to W that "preserves" products. Specifically,

$$T(vw) = T(v)T(w) \quad \text{for all } v, w \text{ in } V.$$

An algebra isomorphism is an algebra homomorphism with an inverse function. It is not hard to show that this inverse function must be linear also, and an algebra homomorphism. So the inverse function is an algebra isomorphism too, from W to V.

Two algebras are called algebra isomorphic if there is an algebra isomorphism between them. And algebra isomorphic algebras are thought of as identical in all important respects: they are not really different algebras at all, but merely two manifestations of the same structure.

The Frobenius theorem, proved in 1877, asserts that there are exactly three finite dimensional associative real division algebras.

The first is the real numbers, the one dimensional example. The second is the complex numbers, a two dimensional algebra, about which we will meditate next. The third is a four dimensional algebra, the main subject of these notes, the quaternions.

This theorem asserts, more specifically, that *any* finite dimensional associative real division algebras is algebra isomorphic to one of these three.

An **algebra norm** on a real algebra V is a norm on V that satisfies a condition involving the product. Specifically, a non-negative real valued function $\|\cdot\|$ defined on V is an algebra norm if, for every real number r and any vectors v and w

$$\begin{aligned}
&\| r\, v \| = |r|\, \|v\| \\
&\| v + w \| \leq \|v\| + \|w\| \quad \text{(the triangle inequality)} \\
&\|v\| = 0 \;\text{ if and only if }\; v = 0 \;\text{ and} \\
&\| v\, w \| = \|v\|\, \|w\|.
\end{aligned}$$

Hurwitz's theorem, proved in 1898, takes a slightly different tack from the Frobenius theorem. It states that any unital division algebra with an algebra norm is algebra isomorphic to one of the three algebras listed above or to an eight dimensional algebra called either the octonions or, in some older sources, the Cayley numbers. The octonions are not associative, and we will make only the most superficial exploration of their properties in these notes.

The differences between these two theorems, which pin down the possible division algebras to a short list, are that the Frobenius theorem assumes the algebra to be unital and associative and of finite dimension, but no norm is involved. Hurwitz's theorem, on the other hand, assumes neither finite dimensionality nor associativity, but does use a norm and also assumes the existence of a multiplicative identity.

2. Isometries and Linearity in $\mathbb{R}^n$

An **isometry** on $\mathbb{R}^n$ is a function from $\mathbb{R}^n$ to itself that preserves the Euclidean length on $\mathbb{R}^n$. Specifically, H is an isometry exactly when

$$\|H(u) - H(v)\| = \|u - v\| \quad \text{for all } u, v \text{ in } \mathbb{R}^n$$

where the norm indicated here is defined on vector $v = (v_1, v_2, \ldots, v_n)$ by

$$\|\,(v_1, v_2, \ldots, v_n)\,\| = \sqrt{v_1^2 + \cdots + v_n^2}.$$

So an isometry H is uniformly continuous in the sense that for all $\varepsilon > 0$ and for every pair of vectors u and v you can guarantee $\|H(v) - H(u)\| < \varepsilon$ by requiring that $\|v - u\| < \varepsilon$.

We will be interested particularly in isometries for which $H(0) = 0$, and will require this of the isometries we consider, in which case (set $v = 0$) we have

$$\|H(u)\| = \|u\| \quad \text{for all vectors } u.$$

For any u, an application of the triangle inequality gives

$$2\|u\| = \|u - (-u)\| = \|H(u) - H(-u)\| \leq \|H(u)\| + \|H(-u)\| = 2\,\|u\|$$

so equality holds above, which can only happen if $H(u)$ is a positive multiple of $-H(-u)$. Since their norms are equal, this means

$$-H(u) = H(-u) \qquad \text{for all } u.$$

Suppose n is a positive integer and $H(nu) = nH(u)$. Then

$$\begin{aligned}(n+1)\|u\| =& \|H((n+1)u)\| = \|H((n+1)u) - H(nu) + H(nu)\|\\ &\leq \|H((n+1)u) - H(nu)\| + \|H(nu)\| = \|(n+1)u - nu\| + \|H(nu)\|\\ &= \|u\| + \|nu\| = (n+1)\|u\|.\end{aligned}$$

Again, we see equality holds in this application of the triangle inequality, so $H(nu)$ is a positive multiple of $H((n+1)u) - H(nu)$ and in view of their norms, that multiple is n. So

$$n(\,H((n+1)u) - H(nu)\,) = H(nu)$$

which implies $n\,H((n+1)u) = (n+1)\,H(nu)$. Since by our initial assumption $n\,H(u) = H(nu)$ we find $H((n+1)u) = (n+1)\,H(u)$.

Since the assumed condition is obviously true for $n = 1$ an appeal to induction allows us to deduce that the condition we assumed above is true for any u and any positive integer n. Because $H(0) = 0$ and $H(-u) = -H(u)$, the condition $n\,H(u) = H(nu)$ holds for any u and any integer n.

With this in hand, if n is a nonzero integer and p any integer

$$H\left(\frac{p}{n}\,u\right) = p\,H\left(\frac{1}{n}\,u\right) = \frac{p}{n}\,n\,H\left(\frac{1}{n}\,u\right) = \frac{p}{n}\,H\left(n\,\frac{1}{n}\,u\right) = \frac{p}{n}\,H\,(u)\,.$$

Any real number c is the limit of a sequence of rational numbers r_n for integers $n > 1$ and so for vector u the vector cu is the limit of the vector sequence

$$\lim_{n\to\infty} r_n u = cu \quad \text{and also} \quad \lim_{n\to\infty} r_n H(u) = cH(u).$$

Since H is continuous at cu, for any real number c and any vector u

$$cH(u) = \lim_{n\to\infty} r_n H(u) = \lim_{n\to\infty} H(r_n u) = H(cu).$$

Finally, if u and v are vectors

$$\begin{aligned} 2\|u-v\| &= \|2u-2v\| = \|H(2u)-H(2v)\| \\ &= \|H(2u)-H(u+v)+H(u+v)-H(2v)\| \\ &\le \|H(2u)-H(u+v)\| + \|H(u+v)-H(2v)\| \\ &= \|2u-u-v\| + \|u+v-2v\| = 2\|u-v\| \end{aligned}$$

so, as above, we have equality throughout which means

$$H(2u) - H(u+v) = H(u+v) - H(2v)$$

so $H(2u) + H(2v) = 2H(u+v)$. Since $H(2u) = 2H(u)$ and $H(2v) = 2H(v)$ we find

$$H(u) + H(v) = H(u+v) \quad \text{for all vectors } u, v.$$

Our conclusion is that H is a linear function. And since $H(v) = 0$ only when $v = 0$ this function has trivial kernel. Since $\mathbb{R}^n$ is finite dimensional, this means H is onto $\mathbb{R}^n$, and so an isomorphism.

Isometries on $\mathbb{R}^n$ which fix the origin are linear isomorphisms.

The main result (linearity) is a special case of a more general result called the **Mazur-Ulam theorem** which tells us that any function from one real normed vector space *onto* another (finite dimensional or not) which sends the origin to the origin and which preserves distances, as our H did above, must be linear.[2]

3. Isometries and Reflections in $\mathbb{R}^n$

If u is a vector in $\mathbb{R}^n$ define $\mathbf{u}^\perp$ to be the subspace

$$u^\perp = \{\, v \in \mathbb{R}^n \mid u \cdot v = 0 \,\}.$$

This is the collection of all vectors perpendicular to u. If u is nonzero, $u^\perp$ has dimension $n-1$. If $n = 1$ then $u^\perp = \{0\}$. If $n = 2$ then $u^\perp$ is the line of vectors through the origin perpendicular to u. If $n = 3$ then $u^\perp$ is a plane and if $n \ge 4$ it is called a hyperplane.

Now define for nonzero u the linear functions $\mathbf{Proj_{u^\perp}}$ and $\mathbf{Proj_u}$ by

$$Proj_u(v) = \frac{v\cdot u}{u\cdot u}\,u \quad \text{and} \quad Proj_{u^\perp}(v) = v - Proj_u(v).$$

[2] A. Vogt, Maps which preserve equality of distance, Studia Math. 45 (1973) 4348.

These functions are called **projections** onto the line through u and onto the subspace $u^{\perp}$, respectively.

The identity $v = Proj_{u^{\perp}}(v) + Proj_u(v)$ decomposes v into the sum of two vectors, one a multiple of u and the other perpendicular to u.

The function $\mathbf{R_u}$ given by

$$R_u(v) = Proj_{u^{\perp}}(v) - Proj_u(v) = v - 2Proj_u(v)$$

is called the **reflection** in $u_{\perp}$. All vectors in $u_{\perp}$ are left alone by R_u, while any multiple cu of u is sent to $-cu$.

$$\begin{aligned} R_u(v) \cdot R_u(v) &= \left(Proj_{u^{\perp}}(v) - Proj_u(v)\right) \cdot \left(Proj_{u^{\perp}}(v) - Proj_u(v)\right) \\ &= \left(Proj_{u^{\perp}}(v)\right) \cdot \left(Proj_{u^{\perp}}(v)\right) + \left(Proj_u(v)\right) \cdot \left(Proj_u(v)\right) \\ &= \left(Proj_{u^{\perp}}(v) + Proj_u(v)\right) \cdot \left(Proj_{u^{\perp}}(v) + Proj_u(v)\right) \;=\; v \cdot v \end{aligned}$$

and the linearity of R_u then implies that it is an isometry.

The norm of $u = (u_1, \ldots, u_n)$ is irrelevant in the formation of R_u, so let's replace u with $u/\|u\|$ so we can work with unit vectors. In that case, if e_i is the unit vector in the direction of the ith coordinate axis, and if I_n is the $n \times n$ identity matrix, the matrix M of R_u (vectors thought of as columns, matrix on the left) has ith column $R_u(e_i) = e_i - 2u_i u$. So

$$M = I_n - 2uu^T$$

where u^T is the row matrix that is the transpose of column u.

Since R_u is an isometry the columns of M form an orthonormal basis for $\mathbb{R}^n$ and $M^{-1} = M^T$: that is, M is an orthogonal matrix.

M acts as the identity on the subspace $u^{\perp}$. Select $p_1 = u$ and let $p_2, \ldots, p_n$ be an orthonormal basis of $u^{\perp}$. Then

$$p_1, p_2, \ldots, p_n$$

constitutes an orthonormal basis of $\mathbb{R}^n$. Let P be the matrix whose columns are the members of this basis, in order. So

$$M\,p_1 = -p_1 \quad \text{and} \quad M\,p_i = p_i \text{ for } i = 2, \ldots, n$$

and therefore

$$P^T\,M\,P = \begin{pmatrix} -1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

Note that the determinant of this matrix, and so the determinant of M, is -1. The columns of M form a left-handed rather than a right-handed orthonormal basis. (These two groups of bases are determined by the determinant condition, and except in dimension three the visual metaphor associated with "handedness" is not very useful.)

We have discovered that any reflection matrix can be brought to the form on the right above by an orthogonal matrix of transition P with the normalized vector u in the far left column and the remaining columns spanning $u^{\perp}$. On the other hand, any matrix M that can be brought to this form by orthogonal P corresponds to the reflection in $p_1^\perp$ given explicitly by

$$R_{p_1}(v) = (-p_1 \cdot v)\ p_1 + (\ (p_2 \cdot v)\ p_2 + \cdots + (p_n \cdot v)\ p_n\ ).$$

We will now discuss a result often referred to as (a special case of) the **Cartan-Dieudonné theorem.**

The theorem we will prove states that every (nonidentity) linear isometry of $\mathbb{R}^n$ can be formed as a composition of at most n reflections of the type we considered above.

This result is obviously true when $n = 1$, because in that case the only nontrivial linear isometry is the function $H(x) = -x$ which is $R_1(x)$ for any real x.

In dimension 2 the situation is a bit different. There, the matrix of a linear isometry is of the form

$$M = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \quad \text{where} \quad a_1^2 + a_2^2 = 1 = b_1^2 + b_2^2 \quad \text{and} \quad a_1 b_1 = -a_2 b_2.$$

This corresponds to

$$\begin{pmatrix} a & \mp\sqrt{1-a^2} \\ \pm\sqrt{1-a^2} & a \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} a & \pm\sqrt{1-a^2} \\ \pm\sqrt{1-a^2} & -a \end{pmatrix} \quad \text{with} \quad -1 \le a \le 1$$

and where the signs are chosen so that the columns are orthogonal. Matrices of the first type have determinant 1, while those of the second type have determinant -1.

There are unique θ and μ with

$$0 \le \theta < 2\,\pi \quad \text{and} \quad 0 \le \mu < 2\,\pi$$

so that matrices of the first and second types can be represented, respectively, as

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \cos(\mu) & \sin(\mu) \\ \sin(\mu) & -\cos(\mu) \end{pmatrix}.$$

Matrices of the first kind correspond to rotations, counterclockwise by angle θ. Matrices of the second kind correspond to reflections across the line through the origin at angle $\mu/2$, as can most easily be checked by calculating that

$$\begin{pmatrix} \cos(\mu) & \sin(\mu) \\ \sin(\mu) & -\cos(\mu) \end{pmatrix} \begin{pmatrix} \cos\left(\frac{\mu}{2}\right) \\ \sin\left(\frac{\mu}{2}\right) \end{pmatrix} = \begin{pmatrix} \cos\left(\frac{\mu}{2}\right) \\ \sin\left(\frac{\mu}{2}\right) \end{pmatrix}$$

$$\text{and} \quad \begin{pmatrix} \cos(\mu) & \sin(\mu) \\ \sin(\mu) & -\cos(\mu) \end{pmatrix} \begin{pmatrix} -\sin\left(\frac{\mu}{2}\right) \\ \cos\left(\frac{\mu}{2}\right) \end{pmatrix} = \begin{pmatrix} \sin\left(\frac{\mu}{2}\right) \\ -\cos\left(\frac{\mu}{2}\right) \end{pmatrix}$$

There are a number of relationships among these types of orthogonal matrices, but just one that concerns us now.

$$\begin{aligned} &\begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ \sin(\alpha) & -\cos(\alpha) \end{pmatrix} \begin{pmatrix} \cos(\beta) & \sin(\beta) \\ \sin(\beta) & -\cos(\beta) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta) & \cos(\alpha)\sin(\beta) - \sin(\alpha)\cos(\beta) \\ \sin(\alpha)\cos(\beta) - \cos(\alpha)\sin(\beta) & \sin(\alpha)\sin(\beta) + \cos(\alpha)\cos(\beta) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\alpha-\beta) & -\sin(\alpha-\beta) \\ \sin(\alpha-\beta) & \cos(\alpha-\beta) \end{pmatrix}. \end{aligned}$$

So products of two reflection matrices can produce any rotation matrix, in many ways. The case of $\beta = 0$ provides a particularly easy route to finding such a product for a given rotation angle α: first reflect across the x axis, then across the line at angle $\alpha/2$.

Our conclusion is that in $\mathbb{R}^2$ any isometry is either a reflection or the product of two reflections.

Let us suppose we have shown for some k and all n with $0 < n \leq k$ that any isometry in dimension n is the composition of at most n reflections. The case $k = 2$ corresponds to the calculations we have just completed.

Suppose further that we are in possession of a linear isometry H on $\mathbb{R}^{k+1}$. The matrix M of this isometry is orthogonal and therefore normal (it commutes with its transpose) and real so it can be reduced using orthogonal matrix of transition P to block diagonal form

$$P^T M P = \begin{pmatrix} \mathbf{A} & O \\ O & \mathbf{B} \end{pmatrix}$$

where A is either 1×1 (i.e. a number) which must be ± 1 or it is 2×2 in which case it has a conjugate pair of complex eigenvalues of norm 1 and is of the form, for some angle θ,

$$\mathbf{A} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

By assumption the matrix $\mathbf{B}$ is the product of no more than j reflection matrices where j is the width of $\mathbf{B}$, either k or $k-1$.

We consider three cases.

First, if $\mathbf{A}$ is the number 1 then $P^T M P$ is the product of no more than k reflections of the form

$$\begin{pmatrix} 1 & O \\ O & R \end{pmatrix}$$

where R is one of the reflection matrices whose product is $\mathbf{B}$.

Second, if $\mathbf{A}$ is the number -1 then $P^T M P$ is the product of no more than k reflections of the form just described plus the reflection

$$\begin{pmatrix} -1 & O \\ O & I_k \end{pmatrix}$$

where I_k is the $k \times k$ identity matrix, for a total of at most $k+1$ reflection matrices needed to form $P^T M P$.

Finally, if $\mathbf{A}$ is 2×2 we have calculated that $\mathbf{A}$ can be written as the product of 2 reflection matrices and we assume that the $(k-1) \times (k-1)$ matrix $\mathbf{B}$ can be written as the product of at most $k-1$ reflection matrices. We conclude that $P^T M P$ itself can be written as a product of no more than $k+1$ reflection matrices of the form

$$\begin{pmatrix} I_2 & O \\ O & R \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} S & O \\ O & I_{k-1} \end{pmatrix}$$

where S is a 2×2 reflection matrix used to create $\mathbf{A}$.

If $R_1 R_2 \cdots R_m$ is a product of reflections equal to $P^T M P$ then

$$P R_1 R_2 \cdots R_m P^T = \left(P R_1 P^T\right)\left(P R_2 P^T\right) \cdots \left(P R_m P^T\right)$$

is a product of reflections equal to M. We have just shown that at most $k+1$ reflection matrices are required for our matrix M, and invoke induction to conclude that this is true not just for *this* k but for *any* k, and the theorem is proved.

As pointed out by Conway and Smith[3] there is a little more that can be squeezed out of this argument. If a linear isometry fixes not just the origin, but acts as the identity on any m dimensional subspace of $\mathbb{R}^n$ for $0 \leq m < n$ then it can be formed as the composition of at most $n-m$ reflections.

Any isometry on $\mathbb{R}^n$ which acts as the identity on any m dimensional subspace of $\mathbb{R}^n$ for some m with $0 \leq m < n$ can be realized as the composition of at most $n-m$ reflections.

We note also that if the isometry has determinant 1 there must be an even number of reflections involved in the product. If the determinant is -1 there are an odd number of reflections involved.

4. (*) Complex Numbers: First Steps

Quaternions are sometimes referred to as "hypercomplex numbers" and the story of quaternions, their geometry and applications and properties, is tied up with the story of complex numbers so we will start there.

Complex numbers are usually introduced to algebra students, around the time they learn about the quadratic formula, as "expressions of the form"

$$a + b\,i$$

where a and b are real numbers and i is a mysterious symbol having the property

$$i^2 = -1$$

and which is otherwise handled formally just like one would deal with a variable in a polynomial expression.

The symbol $\mathbb{C}$ is used to denote the set of complex numbers.

The real numbers a and b are called the **real and imaginary parts** of the complex number, respectively. If a is 0 the number is called **imaginary** while if b is 0 the complex number is called **real.**

In attempts to solve polynomial equations, expressions involving quantities with negative squares appeared in European mathematics in the mid-sixteenth century (Cardano and Bombelli) but were widely regarded as serving only intermediate purpose, en route to true solutions which had to be (positive, at first) real numbers.

More than 200 years later Euler, Argand and (ultimately) Gauss developed the link between planar geometry and complex numbers in more or less its present form, and that is the story of the next sections. They learned that i, though not a real number, corresponds to something very real. It can represent a $90°$ counterclockwise rotation in the plane, and the equation $i^2 = -1$ means that $(1,0)$ is rotated onto $(-1,0)$ by two such rotations.

---

[3] **John Conway** and **Derek Smith**: On Quaternions and Octonions. A. K. Peters, Ltd. 2003, p. 6. This is a superb and very readable little book, recommended for anyone who wants to delve a little deeper into the properties of these division algebras.

We will also use the vocabulary of vector spaces and linear functions, already encountered in Section 1, which are a more recent invention, the definitions of which were organized and clarified by Peano in 1888.

The first, algebraic, definition of complex numbers is as the collection of symbols of the form given above together with two operations, addition and multiplication, codified by the formulae

$$(a + b\,i) + (c + d\,i) = (a + b) + (c + d)\,i$$

and

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

for real a, b, c and d.

It is hardly obvious by cursory inspection that various common properties of the ordinary operations on numbers, such as the distributive laws, associativity and commutativity, actually hold for these operations. Are there multiplicative and additive identities? Are there multiplicative and additive inverses?

It is a calculation to show that all the common laws do in fact hold and that $0 = 0 + 0\,i$ is the additive identity and $1 = 1 + 0\,i$ is the multiplicative identity. The number $a + b\,i$ has additive inverse $-a + (-b)\,i$.

The **complex conjugate** of $a + b\,i$ is denoted $\overline{\mathbf{a} + \mathbf{b}\,\mathbf{i}}$ and defined to be $a + (-b)\,i = a - b\,i$. One shows that for complex numbers z and w that

$$\overline{z + w} = \overline{z} + \overline{w} \quad \text{and} \quad \overline{zw} = (\overline{z})\,(\overline{w}).$$

Note that $(a + b\,i)(\overline{a + b\,i}) = a^2 + b^2$ so unless both a and b are 0 we find

$$1 = (a + b\,i)\left(\frac{a}{a^2 + b^2} + \frac{-b}{a^2 + b^2}\,i\right) = (a + b\,i)\left(\frac{\overline{a + b\,i}}{(a + b\,i)\,(\overline{a + b\,i})}\right).$$

The calculation shows that nonzero complex number $z = a + b\,i$ has multiplicative inverse given by

$$z^{-1} = \frac{\overline{z}}{z\,\overline{z}}\ .$$

Possessing the requisite properties, $\mathbb{C}$ is a two dimensional commutative associative real division algebra.

The **norm** of $a + b\,i$ is denoted $\|\mathbf{a} + \mathbf{b}\,\mathbf{i}\|$ and defined to be

$$\|a + b\,i\| = \sqrt{a^2 + b^2} = \sqrt{z\,\overline{z}}.$$

If r is real it is obvious that $\|rz\| = |r|\,\|z\|$, but even more is true. Actually,

$$\|zw\| \;=\|z\|\|w\| \quad \text{for any complex numbers } z \text{ and } w.$$

Also

$$\|z + w\| \;\le\; \|z\| + \|w\| \quad \text{(the triangle inequality)}$$
$$\text{and} \quad \|z\| \ge 0 \text{ with equality when and only when } z = 0.$$

Collectively, these contain the required qualities of an algebra norm on an algebra, in this particular case a norm on $\mathbb{C}$.

Finally, we introduce what is known as the **polar form** of complex numbers. A calculation shows that for any nonzero complex number $z = a + b\,i$, the complex number $z/\|z\|$ has norm 1. It's coefficients are

$$a/\sqrt{a^2+b^2} \quad \text{and} \quad b/\sqrt{a^2+b^2}$$

and there is exactly one angle θ in $(-\pi, \pi]$, sometimes denoted $\mathbf{arg(z)}$ and called the **argument** of z with

$$\cos(\theta) = a/\sqrt{a^2+b^2} \quad \text{and} \quad \sin(\theta) = b/\sqrt{a^2+b^2}.$$

If we let $r = \|z\|$ then

$$z = a + b\,i = \|z\|\frac{z}{\|z\|} = r\,(cos(\theta) + i\,\sin(\theta))$$

relating the "$a+b\,i$" form of z, also called the rectangular form, with this new form called the polar form which, of course, looks a lot like polar coordinates.

5. Complex Power Series

If the norm of a sequence of complex numbers converges to 0, both the real and imaginary parts of the complex number must converge to zero. The converse also is true.

Properties of the norm bring concepts of convergence, introduced for real sequences, into the complex universe. A complex sequence converges to a complex number if the norm of the sequence of differences converges to 0. This happens exactly when *both* real and imaginary parts of the sequence of differences converge to 0.

Since a series is just a type of sequence, we also know when a series, and in particular a **power series**, formed from complex numbers will converge.

In Calculus we show that the three series

$$1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \frac{x^6}{720} \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!} \;=\; e^x$$

and

$$1 - \frac{x^2}{2} + \frac{x^4}{24} - \frac{x^6}{720} + \cdots = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} \;=\; \cos(x)$$

and

$$x - \frac{x^3}{6} + \frac{x^5}{120} - \cdots = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!} \;=\; \sin(x)$$

all converge with infinite radius of convergence, and if you wish to proceed that way could constitute the *definition* of the indicated functions.

A series of this type, which is the limit of a specific sequences of partial sums, can be reorganized and broken into "sub-series" whose terms are listed in any order we like (so long as all terms are eventually included) and the limit will be unchanged.

Standard results on product series, the key to one proof that

$$e^x e^y = e^{x+y},$$

rely on the real norm: i.e. the absolute value.

When we use the complex norm to define convergence of complex series, judicious application of the triangle inequality gives these same results for complex series.

Any real function with a power series expansion can be extended to a complex domain, and the domain of convergence will contain all complex numbers with norm less than the real radius of convergence.

Let me reiterate this amazing fact: most real functions we encounter have power series with, at least, *some* radius of convergence. All of these can be "complexified" to complex domain, producing complex output.

Euler's equation, seen below, is an example of this in action.

$$\mathbf{e^{a+b\,i}} = e^a e^{b\,i} = e^a\left(\cos(b) + i\,\sin(b)\right).$$

Note that the right hand side is the polar form of any complex number that can be written as $z = e^{a+b\,i}$. That includes any complex number except 0.

The formula is proved by substituting $b\,i$ into the power series for e^x and algebraically simplifying the powers $(b\,i)^n$. Gathering terms in which a factor of i remains and factoring out this common factor produces the series for $\sin(b)$. The remaining terms are the series for $\cos(b)$.

If $e^{a+b\,i} = e^{c+d\,i}$ for real a, b, c and d then $a = c$ and $b - d$ is a multiple of 2π. So the complex exponential has no global inverse function. When its domain is restricted there can be inverses. In particular, when restricted to complex numbers $a + b\,i$ where a is real and b is taken from an interval of the form $(h,\, h + 2\pi)$ the exponential is one-to-one and therefore has an inverse on its range. That range consists of all nonzero complex numbers whose polar form uses an angle *not* coterminal with h.

The inverse on this particular piece of the exponential function is usually denoted with some form of **logarithm** notation such as

$$\log(\mathbf{z}) = \ln(\|z\|) + i\,\theta$$

when $z \neq 0$ and θ is the unique angle, $\mathbf{arg(z)}$, in the interval $(h, h + 2\pi)$ for which $z = \|z\|\,(\cos(\theta) + i\,\sin(\theta))$.

An important ambiguity arises here, caused by the fact that the exponential has no global inverse. $arg(z)$, of course, depends on which piece of the exponential we are using, and that must be specified. Each different h produces what is called a branch of the logarithm.

If it is not specifically mentioned, it is the custom to choose $h = -\pi$ so that the excluded complex numbers in the range of the exponential are those which lie along the negative real axis, and $arg(z)$ is taken from $(-\pi, \pi)$.

Given a particular branch of the logarithm, we can define for complex numbers p and q (with p *not* excluded from the domain of this branch)

$$\mathbf{p^q} = e^{q\,\log(p)}.$$

Other important facts are

$$\begin{aligned}&(\cos(b) + i\sin(b))\,(\cos(c) + i\sin(c))\\&\qquad = e^{ib}\,e^{ic} = e^{i(b+c)} = (\cos(b+c) + i\sin(b+c))\end{aligned}$$

and **De Moivre's formula**: for any integer n

$$(\cos(b) + i\sin(b))^n = \left(e^{ib}\right)^n = e^{inb} = (\cos(nb) + i\sin(nb))\,.$$

Functions of a complex variable can have derivatives, defined very similarly to the derivative of an ordinary real function.

If f has complex domain

$$\mathbf{f}'(\mathbf{z}) = \lim_{\|h\|\to 0} \frac{f(z+h) - f(z)}{h} \quad \text{wherever the limit exists.}$$

The **derivative** is the complex number to which the ratio is close, provided only that the norm of h is small.

The usual differentiation formulas hold and if a function is presented as a power series the derivative can be obtained by term-by-term differentiation on the interior of the domain of convergence, just as with real power series, and the proof is essentially identical to the proof for real series.

The study of complex integrals and derivatives is beautiful and very useful to practical folk such as engineers, who spend quite a bit of time with these operations.

The existence of a complex derivative on and around a point in the complex plane has strong and surprising implications, whose explication lies beyond the boundary of these notes.

6. (*) The Geometry of Complex Numbers

Given the basic facts about complex numbers we make a new description. Vectors in the plane can be defined by two real numbers. So can complex numbers. So we can make an association between vectors in the plane and complex numbers.

Complex numbers are algebraic objects, characterized by the unusual complex multiplication. Vectors in the plane are all about geometry and angles and lengths.

The association will be useful if complex multiplication implies something interesting about geometry, or if geometry will help us understand complex multiplication. In fact, both occur.

The most obvious association is the one we pick.

Identify complex number $z = a + b\,i$ with the vector $\langle a, b\rangle$.

Note that norm in the plane, the pythagorean distance from tip to tail of a vector, agrees with the norm of the associated complex number.

Recall that if the *point* (a, b) has polar coordinates (r, θ) then the *vector* $\langle a, b\rangle$ can be written as

$$\langle a, b\rangle = r\mathbf{U}_\theta$$

where $\mathbf{U}_\theta = \langle \cos(\theta),\, \sin(\theta) \rangle$ is the vector with its nose resting on the unit circle at angle θ, pointing the way, and $r = \sqrt{a^2 + b^2}$ is a "stretch" (or "shrink") factor.

Vectors are entirely concerned with norm and direction, and the rectangular coordinates of a vector don't display those qualities in the most convenient way.

This **polar form** of the vector *does*, and is extremely useful in applications of vectors. After all, we wouldn't be *using* vectors if we weren't obsessed with these features, and in most applications that really *use* vectors we move back and forth between rectangular form, convenient for combining vectors, and polar form, emphasizing their norm and direction.

$$\langle a, b \rangle \quad \Longleftrightarrow \quad r\,\mathbf{U}_\theta = r\,\langle \cos(\theta),\, \sin(\theta) \rangle = \langle r\cos(\theta),\, r\sin(\theta) \rangle\,.$$

With our association between vectors in the plane and complex numbers, the complex number $e^{i\theta} = \cos(\theta) + i\,\sin(\theta)$ is associated with unit vector $\mathbf{U}_\theta$ and, generally,

$$a + b\,i = r\,e^{i\theta} \quad \Longleftrightarrow \quad \langle a,\, b \rangle = r\,\mathbf{U}_\theta = \langle r\cos(\theta),\, r\sin(\theta) \rangle\,.$$

We now can use complex multiplication to simplify our thinking about rotations and dilations in the plane.

The calculations from above show that multiplying a complex number z by $r\,e^{i\theta}$ (where r is positive, θ real) has the effect of changing its norm by factor r and rotating it by angle θ.

$$\text{If } z = s\,e^{i\mu} \text{ then } (r\,e^{i\theta})z = rs\,e^{i(\theta+\mu)}.$$

If we want to rotate a vector $\langle a,\, b \rangle$ ("tail", of course, at the origin) by θ followed by a 15% stretch, calculate

$$1.15\,e^{i\theta}\,(a + b\,i) = 1.15\,(\cos(\theta) + i\,\sin(\theta))\,(a + b\,i)\,.$$

The real part of this product is the first component of the stretched, rotated vector and the imaginary part is the second component.

If we have several consecutive rotations we can combine them into a single effective rotation by multiplying the complex numbers corresponding to the constituent rotations, and the result will be a single complex number corresponding to the final rotation.

7. Matrix Form of Complex Numbers

Since rotations and dilations in the plane are linear transformations, and all linear transformations on $\mathbb{R}^2$ are given by left matrix multiplication, this action of complex numbers on vectors can be given by matrix multiplication.

The matrix associated with complex number $a + b\,i = r\,e^{i\theta}$ is

$$a + b\,i \quad \Longleftrightarrow \quad r\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Matrices generally do not commute, but *these* matrices do commute with each other, and a calculation shows that matrix multiplication corresponds to complex multiplication. Note that the square of the second matrix is the negative of the identity matrix, so the second matrix "acts like" i, while the first matrix "acts like" 1.

And the squared norm of the associated complex number is the determinant of the matrix of that complex number.

When the matrix has norm 1 (i.e. $r = 1$) this matrix is simply the matrix of a rotation in the plane.

So everything about complex numbers can be expressed in terms of these matrices. Complex conjugation, for example, corresponds to taking the transpose.

This two dimensional real vector space of matrices with matrix multiplication is algebra isomorphic to the complex numbers.

The multiple representations of these complex numbers, the original representation, vectors and finally matrices with matrix operations, are all consistent. For instance

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax - by \\ bx + ay \end{pmatrix} \Leftrightarrow \begin{pmatrix} ax - by & -bx - ay \\ bx + ay & ax - by \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

and, of course

$$(a + b\,i)(x + y\,i) = (ax - by) + (bx + ay)\,i \Longleftrightarrow \begin{pmatrix} ax - by & -bx - ay \\ bx + ay & ax - by \end{pmatrix}.$$

To recapitulate, vectors in the plane, this set of matrices and the complex numbers as originally presented are each two dimensional real vector spaces. The commutative, associative operation provided by matrix multiplication (i.e. complex multiplication) has multiplicative identity and every nonzero element has a multiplicative inverse. The distributive properties hold. The given association between the three forms reveal them to be algebra isomorphic.

And there is a norm, giving us an idea of closeness so ideas from calculus can be extended to complex functions if we want.

8. (*) Linear Rigid Motions in Space

An ordered basis $\vec{u}_1$, $\vec{u}_2$, $\vec{u}_3$ of $\mathbb{R}^3$ is called **right-handed** if the determinant

$$\det\left(\vec{u}_1\, \vec{u}_2\, \vec{u}_3\right) > 0,$$

where the coordinates of the basis vectors in terms of the standard basis are given as columns. It is called **left-handed** if that determinant is negative.

We define here a **linear rigid motion** $\boldsymbol{\mathcal{L}}$ on $\mathbb{R}^3$ to be a linear map that sends any (and hence every) right-handed ordered orthonormal basis, sometimes called an orthonormal frame, to another. In particular it must send the **standard basis vectors**

$$\vec{\boldsymbol{i}} = \langle 1, 0, 0 \rangle\,, \quad \vec{\boldsymbol{j}} = \langle 0, 1, 0 \rangle\,, \quad \vec{\boldsymbol{k}} = \langle 0, 0, 1 \rangle$$

to an orthonormal frame $\vec{u}_1$, $\vec{u}_2$, $\vec{u}_3$ and the matrix $\mathbf{M} = \left(\vec{u}_1\, \vec{u}_2\, \vec{u}_3\right)$ must have determinant one (it is the volume of a unit cube.)

A linear map of this kind must be an isometry; that is, it cannot change the length of any vector. So any real eigenvalue must have value ± 1 and any complex eigenvalues have norm 1 and come as a conjugate pair.

$\mathbf{M}$ is the matrix of this linear rigid motion: that is,

$$\boldsymbol{\mathcal{L}}(v) = \mathbf{M}v \quad \text{for every vector } v \text{ in } \mathbb{R}^3$$

and $\mathbf{M}$ has characteristic polynomial

$$\lambda^3 + a\lambda^2 + b\lambda - \det \mathbf{M} = \lambda^3 + a\lambda^2 + b\lambda - 1$$

for certain real numbers a and b.

Since the characteristic polynomial has third degree, this matrix has at least one real eigenvalue.

As a first case, if our real eigenvalue is -1 then $\lambda + 1$ divides the characteristic polynomial, leaving quotient of the form

$$\lambda^2 + c\lambda - 1$$

which has two real roots, one positive and one negative. Our conclusion: -1 is a root repeated twice, and 1 is a root once.

So $\boldsymbol{\mathcal{L}}$ is a 180° rotation around the axis of an eigenvector with eigenvalue 1, in the plane of eigenvectors with eigenvalue -1.

In the second case we have eigenvalue 1, and after dividing the characteristic polynomial by $\lambda - 1$ we have quotient of the form

$$\lambda^2 + c\lambda + 1.$$

We could have $c = 2$ in which case we have eigenvalue -1 repeated twice and we are in the situation of the first case.

We could have $c = -2$ in which case we have eigenvalue 1 repeated twice more and $\boldsymbol{\mathcal{L}}$ is the identity transformation.

And finally, we might have two complex conjugate eigenvalues, each of norm 1, which must therefore have the form

$$\lambda_1 = \cos(\theta) + i\,\sin(\theta) \quad \text{and} \quad \lambda_2 = \cos(\theta) - i\,\sin(\theta)$$

for unique angle θ in the interval $(-\pi, \pi]$.

Appealing to basic facts about real normal matrices ($\mathbf{M}$ commutes with its transpose) there is an orthogonal basis in which $\mathbf{M}$ has the form

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{pmatrix},$$

More specifically, there is a right-handed ordered orthonormal basis $\vec{p}_1$, $\vec{p}_2$, $\vec{p}_3$ so that the orthogonal matrix of transition $\mathbf{P}$ whose columns are this basis (so $\mathbf{P}^{-1}$ is the transpose $\mathbf{P}^T$) satisfies

$$\mathbf{P}^{-1}\mathbf{M}\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

This is a rotation in the plane of the second and third basis vectors, whose axis of rotation is the first basis vector, an eigenvector for eigenvalue 1.

I should point out that we have adopted the point of view that **M** represents the effect of the rotation on vectors, in particular the coordinate axes. The coordinates of points in this new rotated basis will change in a way that is "opposite" to this: i.e. replace θ by $-\theta$ to get a matrix that produces coordinates in the new basis. This will be, in fact, $\mathbf{M}^{-1}$.

If you begin by *knowing* the normalized axis of rotation $\vec{a}$ and angle θ it is fairly straightforward to produce the matrix. Find any unit vector $\vec{p}_2$ perpendicular to $\vec{p}_1 = \vec{a}$ and let $\vec{p}_3 = \vec{p}_1 \times \vec{p}_2$.

If $\mathbf{P}$ has columns $\vec{p}_1, \vec{p}_2, \vec{p}_3$ then the matrix of the rotation is

$$\mathbf{M} = \mathbf{P}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{pmatrix}\mathbf{P}^{-1}.$$

9. (*) Quaternions

In 1775 Euler proved the **Euler rotation theorem**, that any rigid motion of space that leaves at least one point fixed corresponds to a rotation about an axis through that point. So compositions of rotations, one after the other, that leave this point fixed are also rotations. This is tantalizingly similar to the fact that multiplying two complex numbers of norm one produces another, and multiplication by such a number corresponds to a rotation.

Complex numbers deal efficiently with rotations in the plane, so it was thought that some similar operation on ordered triples could be helpful when trying to solve the much more troublesome issue of rotations in space. The search was on for an operation to make $\mathbb{R}^3$ into a associative division algebra, as complex multiplication did for $\mathbb{R}^2$.

In October, 1843 **William Rowan Hamilton** had an inspiration (reportedly while walking by a particular bridge in Dublin) about how this could be done, but using 4 dimensions, not 3.

The **quaternions** are represented, usually, as the collection of symbols of the form

$$a + b\,i + c\,j + d\,k \text{ where } a, b, c \text{ and } d \text{ are real}$$

and the symbols i, j and k satisfy the equations

$$i^2 = j^2 = k^2 = -1 \text{ and } ij = k \text{ and } jk = i \text{ and } ki = j$$

but otherwise expressions are added and multiplied formally, as polynomial expressions would be, with commutative addition and applications of the distributive law of multiplication over addition. Associativity of this multiplication is assumed.

This set of symbols with these operations is usually denoted $\mathbb{H}$, and the multiplication is called the **Hamilton product**.

It was *not* understood that these properties were consistent, that some might not forbid others in the presence of associativity, and it is a bit of a job to prove them to *be* consistent, unlike the case of complex numbers where the matter is pretty straightforward.

The fact that the multiplication values assumed to be true above imply that

$$ij = -ji \text{ and } jk = -kj \text{ and } ki = -ik$$

and that, in general, quaternions do not commute using this Hamilton product, was regarded as troubling but necessary: after all, rotations are not commutative, as a minute spent rotating an apple can confirm.

The job of showing associativity of the Hamilton product (and consistency of the basic "multiplication table" whose entries are listed above) will be easy for us after we find a collection of matrices to represent our quaternions, as we did with complex numbers. Matrix multiplication will correspond to the Hamilton product, and we know that matrix multiplication is associative and satisfies the distributive properties. Until then we will just assume the properties listed above to be consistent.

As we did with complex numbers, quaternions are identified with $\mathbb{R}^4$

$$a + b\,i + c\,j + d\,k \Longleftrightarrow (a, b, c, d).$$

Even more handy will be the representation as

$$a + b\,i + c\,j + d\,k \Longleftrightarrow a + \vec{v} \text{ where } \vec{v} = \langle b, c, d\rangle = b\,\vec{\boldsymbol{i}} + c\,\vec{\boldsymbol{j}} + d\,\vec{\boldsymbol{k}}.$$

Here the number a is called the **real part** of the quaternion, and $\vec{v}$ is called the **imaginary part**.

In the text that follows, the imaginary part of a quaternion is identified with ordinary three dimensional space, and $\vec{\boldsymbol{i}}$, $\vec{\boldsymbol{j}}$ and $\vec{\boldsymbol{k}}$ are the usual unit coordinate vectors in $\mathbb{R}^3$.

If $a = 0$ and $\vec{v} \neq 0$ the quaternion is called **imaginary**, while if $\vec{v} = 0$ the quaternion is called **real**.

Expressed this way, the "multiplication table entries" shown above lead to a formula for the Hamilton product of quaternions $q_1 = a_1 + \vec{v}_1 = a_1 + b_1\vec{\boldsymbol{i}} + c_1\vec{\boldsymbol{j}} + d_1\vec{\boldsymbol{k}}$ on the left by $q_2 = a_2 + \vec{v}_2 = a_2 + b_2\vec{\boldsymbol{i}} + c_2\vec{\boldsymbol{j}} + d_2\vec{\boldsymbol{k}}$ on the right:

$$\begin{aligned} q_1\, q_2 &= (a_1 + \vec{v}_1)\,(a_2 + \vec{v}_2) \\ &= (a_1 a_2 - \vec{v}_1 \cdot \vec{v}_2) + (a_1\vec{v}_2 + a_2\vec{v}_1 + \vec{v}_1 \times \vec{v}_2) \end{aligned}$$

where $\vec{v}_1 \cdot \vec{v}_2$ represents the ordinary real dot product

$$\vec{v}_1 \cdot \vec{v}_2 = b_1 b_2 + c_1 c_2 + d_1 d_2$$

and $\vec{v}_1 \times \vec{v}_2$ denotes cross product

$$\vec{v}_1 \times \vec{v}_2 = (c_1 d_2 - c_2 d_1)\,\vec{\boldsymbol{i}} + (d_1 b_2 - b_1 d_2)\,\vec{\boldsymbol{j}} + (b_1 c_2 - c_1 b_2)\,\vec{\boldsymbol{k}}.$$

It is now easy to see the source of the lack of commutativity in the Hamilton product. If you switch the order of the two quaternions, all the terms in the product remain the same *except* the cross product term. And cross product is **anti-commutative**, i.e.

$$\vec{v}_1 \times \vec{v}_2 = -\vec{v}_2 \times \vec{v}_1.$$

So the real part of the Hamilton product commutes, but the imaginary part does not due to its cross product component. *When and only when the imaginary parts are parallel (or one is zero) will the Hamilton product of two quaternions commute.*

For $q \in \mathbb{H}$ we will sometimes use an "arrow on top" notation such as $q = q_r + \vec{q}_i$ or $q = a + \vec{b}$ to indicate the real and imaginary parts of q, and our intent is that q_r (or a) denote the real part of q and $\vec{q}_i$ (or $\vec{b}$) its imaginary part.

The **conjugate** of $q \in \mathbb{H}$ is denoted $\mathbf{q}^*$ and defined to be $q_r - \vec{q}_i$. This means q^* commutes with q. Also, a calculation shows that

$$(p\,q)^* = q^*\,p^*.$$

If $q = a + b\,\vec{\boldsymbol{i}} + c\,\vec{\boldsymbol{j}} + d\,\vec{\boldsymbol{k}}$ then

$$q\,q^* = q^* q = a^2 + b^2 + c^2 + d^2$$

and $\|q\| = \sqrt{q\,q^*} = \|q^*\|$ is a **norm** on $\mathbb{H}$, satisfying all the properties listed for the complex norm in Section 4.

For instance $\|\,p\,q\,\|^2 = p\,q\,(p\,q)^* = p\,q\,q^* p^* = p\,\|q\|^2\,p^* = \|p\|^2\|q\|^2$.

This means:

$$\|\,p\,q\,\| = \|p\|\,\|q\|.$$

Unless both real and imaginary parts of q are zero

$$\frac{q\,q^*}{\|q\|^2} = q\,\frac{q^*}{\|q\|^2} = q^*\,\frac{q}{\|q\|^2} = 1.$$

So any nonzero q (and q^* too) has a multiplicative inverse given explicitly as above for the Hamilton product.

As in the complex numbers, any nonzero member q of $\mathbb{H}$ can be written in a unique way as a unit quaternion multiplied by a positive scalar:

$$q = \|q\|\,\frac{q}{\|q\|} = \|q\|\,\mathbf{V}_q = \|q\|\left(\frac{q_r}{\|q\|} + \frac{\vec{q}_i}{\|q\|}\right)$$

where the unit quaternion $\mathbf{V}_q$ is called the **versor** of q.

$\vec{q}_i/\|q\|$ will not (usually) be a unit vector, but if it is nonzero this component can be written as

$$\frac{\vec{q}_i}{\|q\|} = \frac{\|\vec{q}_i\|}{\|q\|}\,\frac{\vec{q}_i}{\|\vec{q}_i\|}$$

and we have

$$q = \|q\|\left(\frac{q_r}{\|q\|} + \frac{\vec{q}_i}{\|q\|}\right) = \|q\|\left(\frac{q_r}{\|q\|} + \frac{\|\vec{q}_i\|}{\|q\|}\,\frac{\vec{q}_i}{\|\vec{q}_i\|}\right)$$

where

$$\left(\frac{q_r}{\|q\|}\right)^2 + \left(\frac{\|\vec{q}_i\|}{\|q\|}\right)^2 = 1.$$

Suppose q is *not* real. Then there is an angle θ in $(0,\, \pi)$ and angle $\overline{\theta} = 2\,\pi - \theta$ in $(\pi,\, 2\,\pi)$ for which

$$q = \|q\| \,(\, \cos(\theta) + \sin(\theta)\,\mathbf{U}\,) = \|q\| \left(\, \cos\left(\overline{\theta}\right) + \sin\left(\overline{\theta}\right)(-\mathbf{U})\,\right)$$

where $\mathbf{U}$ is the unit imaginary quaternion $\vec{q}_i/\|q\|$.

This assemblage of factors is called the **polar form** of q, but remember that θ is not determined unless you specify $\vec{q}_i/\|q\|$ or $-\vec{q}_i/\|q\|$ as the unit imaginary.

If q *is* real, a choice of $\theta = \pi$ or $\theta = 0$ produces a polar form; in this case $\mathbf{U}$ can be picked however you like because $\sin(\theta) = 0$.

With the properties assembled in this section (assuming, of course, associativity and consistency which will be dealt with later) we see that $\mathbb{H}$ is a 4 dimensional associative real division algebra. It is not commutative. It has an algebra norm, so we can think of convergence and calculus involving quaternions.

10. (*) Solutions of $z^2 = -1$

There are no solutions among real numbers to $z^2 = -1$, a situation that led to the original manipulations which we would call complex arithmetic.

If we search for solutions among complex numbers $z = a + b\,i$, we see that

$$-1 = a^2 - b^2 + 2abi \Rightarrow a = 0 \text{ and } b^2 = 1.$$

So $b = \pm 1$ and therefore $z = \pm i$. There are exactly two square roots of -1 among complex numbers.

Let's now look for solutions to $q^2 = -1$ among quaternions, and use the polar form $q = \|q\| \,(\, \cos(\theta) + \sin(\theta)\,\mathbf{U}\,)$. So if in fact $q^2 = -1$ then

$$\begin{aligned} -1 =& \|q\|^2 \left(\, \cos^2(\theta) - \sin^2(\theta)\,\mathbf{U}\cdot\mathbf{U} + 2\,\cos(\theta)\sin(\theta)\mathbf{U} + \mathbf{U}\times\mathbf{U}\,\right) \\ =& \|q\|^2 \left(\, \cos^2(\theta) - \sin^2(\theta) + 2\,\cos(\theta)\sin(\theta)\mathbf{U}\,\right). \end{aligned}$$

Since $\cos(\theta)\sin(\theta)$ must be 0 (the imaginary part must vanish) we find that in fact $\cos(\theta) = 0,\ \sin(\theta) = \pm 1$ and $\|q\| = 1$.

So among quaternions solutions can only be found among the unit imaginary quaternions: the vectors on the **unit sphere** $\mathbb{S}^2$ in $\mathbb{R}^3$. And vectors on this sphere are all, in fact, solutions.

Vectors on the unit imaginary sphere are all, and the only,
square roots of -1 among the quaternions.

Examining a quaternion product

$$\begin{aligned} q_1 q_2 &= (a_1 + \vec{v}_1)\,(a_2 + \vec{v}_2) \\ &= (a_1 a_2 - \vec{v}_1 \cdot \vec{v}_2) + (a_1 \vec{v}_2 + a_2 \vec{v}_1 + \vec{v}_1 \times \vec{v}_2) \end{aligned}$$

we see that if $\vec{v}_1$ is a multiple of $\vec{v}_2$ then the product quaternion has imaginary part that is *also* a multiple of $\vec{v}_2$.

Let $\mathbf{U}$ be any unit imaginary. Introducing notation, we define the **slice** of $\mathbb{H}$ containing $\mathbf{U}$ by

$$\mathbb{H}_{\mathbf{U}} = \{\, a + b\,\mathbf{U} \mid a, b \text{ real} \,\} = \mathbb{H}_{(-\mathbf{U})}.$$

$\mathbb{H}_{\mathbf{U}}$ is a two dimensional real vector space, and by the above calculation it is an algebra. The norm on the quaternions gives us a norm on each slice. Each quaternion with nonzero imaginary part is in exactly one of these slices.

The obvious linear map sending complex number $a + b\,i$ to quaternion $a + b\,\mathbf{U}$ is an algebra isomorphism. Also, the map sending $a + b\,i$ to quaternion $a + b\,(-\mathbf{U})$ is an algebra isomorphism. Since an algebra isomorphism must send i to a square root of -1, these are the *only* algebra isomorphisms between $\mathbb{C}$ and $\mathbb{H}_{\mathbf{U}}$.

And the complex norm on $\mathbb{C}$ agrees with the quaternion norm on each $\mathbb{H}_{\mathbf{U}}$.

$\mathbb{H}$ is the union of "copies" of $\mathbb{C}$. These copies are the sets of quaternions $\mathbb{H}_{\mathbf{U}}$, which we called slices above. The real quaternions are in every slice, but those are the only quaternions in more than one slice.

11. (*) Special Products of Quaternions: Rotations in $\mathbb{R}^3$

We have decided to identify vectors in space with imaginary quaternions, and wish to use quaternions to study rotations in space, by analogy with what we did in two dimensions with complex numbers.

There, multiplying by a complex number of norm one did the job.

Exploring this idea, we multiply a quaternion $q = a + \vec{b}$ by a generic nonzero vector $\vec{v}$ which gives

$$q\,\vec{v} = -\vec{b}\cdot\vec{v} + a\vec{v} + \vec{b}\times\vec{v}.$$

If this is to be a vector for any $\vec{v}$, as it must if this Hamilton product is to induce a rotation of $\vec{v}$, then $\vec{b}\cdot\vec{v}$ must equal 0 for any $\vec{v}$, which forces $\vec{b}$ to be the zero vector. So the quaternion q is real. The same is true if we examine the product $\vec{v}\,q$.

If this is to represent rotation we must have $q = a = 1$ so we have discovered the identity "rotation." (If $q = a = -1$ we have **inversion**, which is *not* a rotation in $\mathbb{R}^3$.)

So this attempt fails miserably.

But we shall persist, and examine product $q\,\vec{v}\,p$ involving a *second* quaternion $p = c + \vec{d}$ multiplied on the right and presume, to avoid repeating the last calculation, that neither $\vec{b}$ nor $\vec{d}$ are zero.

Recall the scalar triple product vector identity, for any vectors $\vec{w}$, $\vec{u}$ and $\vec{v}$:

$$\vec{w}\cdot(\vec{u}\times\vec{v}) = \vec{v}\cdot(\vec{w}\times\vec{u}) = \vec{u}\cdot(\vec{v}\times\vec{w})\,.$$

and also the triple cross product identity

$$\vec{w}\times(\vec{u}\times\vec{v}) = (\vec{w}\cdot\vec{v})\;\vec{u} - (\vec{w}\cdot\vec{u})\;\vec{v}.$$

With this in mind, we calculate

$$\begin{aligned} q\,\vec{v}\,p &= \left(-\vec{b}\cdot\vec{v} + a\,\vec{v} + \vec{b}\times\vec{v}\right) p \\ &= -\,c\,\vec{b}\cdot\vec{v} - \left(a\,\vec{v} + \vec{b}\times\vec{v}\right)\cdot\vec{d} + c\left(a\vec{v} + \vec{b}\times\vec{v}\right) - \left(\vec{b}\cdot\vec{v}\right)\vec{d} + \left(a\,\vec{v} + \vec{b}\times\vec{v}\right)\times\vec{d} \\ &= \left(-c\,\vec{b} - a\,\vec{d} + \vec{b}\times\vec{d}\right)\cdot\vec{v} \\ &\qquad + \left(a\,c + \vec{b}\cdot\vec{d}\right)\,\vec{v} - \left(\vec{b}\cdot\vec{v}\right)\,\vec{d} - \left(\vec{d}\cdot\vec{v}\right)\,\vec{b} + \left(c\,\vec{b} - a\,\vec{d}\right)\times\vec{v}. \end{aligned}$$

As above, if multiplication by q and p in this way is to correspond to rotation of generic $\vec{v}$, it must always produce a vector, and so must have zero real part for *any* $\vec{v}$.

$$\left(-c\,\vec{b} - a\,\vec{d} + \vec{b}\times\vec{d}\right)\cdot\vec{v} = 0.$$

Since $\vec{b}\times\vec{d}$ is orthogonal to both $\vec{d}$ and $\vec{b}$ the only way this can happen for *all* $\vec{v}$ is if $\vec{d} = k\vec{b}$ for some constant k, and substitution then yields $c = -k\,a$.

In other words, $p = -k\,\left(a - \vec{b}\right) = -k\,q^*$ for some real constant k.

So we are working with a product of the form

$$-k\,q\,\vec{v}\,q^* = -k\,\|q\|^2\;r\,\vec{v}\,r^*$$

where r is the unit quaternion $q/\|q\|$.

If the norm of $\vec{v}$ is not changed by this multiplication, as it would not be with a rotation, then $-k\,\|q\|^2 = \pm 1$. In other words, we need only consider products as above where q is a *unit* quaternion. But with that assumption, if $-k = -1$ and we determine somehow that $q\,\vec{v}\,q^*$ *does* represent a rotation, then $-1\,q\,\vec{v}\,q^*$ would represent a rotation followed by **inversion**, and in three dimensions inversion is *not* a rotation.

So we will assume that our candidate "rotation" quaternion

$$q = \cos(\theta) + \sin(\theta)\,\mathbf{U} \qquad \text{for} \quad \theta \quad \text{in} \quad (0,\pi)$$

is a unit quaternion in polar form (so $\mathbf{U}$ is a unit imaginary and $\sin(\theta) \geq 0$) and $q\,\vec{v}\,p = q\,\vec{v}\,q^*$, in which case the product becomes

$$\begin{aligned} q\,\vec{v}\,q^* &= \left(a^2 - \vec{b}\cdot\vec{b}\right)\,\vec{v} + 2\,\left(\vec{b}\cdot\vec{v}\right)\,\vec{b} + 2\,a\,\vec{b}\times\vec{v} \\ &= \left(\cos^2(\theta) - \sin^2(\theta)\,\right)\,\vec{v} + \left(2\,\sin^2(\theta)\,\mathbf{U}\cdot\vec{v}\,\right)\,\mathbf{U} + 2\,\cos(\theta)\,\sin(\theta)\,\mathbf{U}\times\vec{v} \\ &= \;\cos(2\,\theta)\,\vec{v} + (\,1 - \cos(2\,\theta))\;(\mathbf{U}\cdot\vec{v}\,)\;\mathbf{U} + \sin(2\,\theta)\,\mathbf{U}\times\vec{v} \end{aligned}$$

We will see that this represents a rotation by angle $2\,\theta$ counterclockwise around axis $\mathbf{U}$.

Suppose $\mathbf{U}, \mathbf{A}$ and $\mathbf{B}$ form an orthonormal ordered basis of $\mathbb{R}^3$, with $\mathbf{U}\times\mathbf{A} = \mathbf{B}$ and $\mathbf{A}\times\mathbf{B} = \mathbf{U}$ and $\mathbf{B}\times\mathbf{U} = \mathbf{A}$. So the basis is a "right-handed" basis, as is the standard basis $\vec{\boldsymbol{i}}$, $\vec{\boldsymbol{j}}$ and $\vec{\boldsymbol{k}}$ of $\mathbb{R}^3$.

Writing $\vec{v}$ in terms of this new basis, we have

$$\vec{v} = x\mathbf{U} + y\mathbf{A} + z\mathbf{B}$$

and then

$$\begin{aligned}
q\,\vec{v}\,q^* =&\ \cos(2\,\theta)\ (x\mathbf{U} + y\mathbf{A} + z\mathbf{B})\\
&\qquad + (\,1 - \cos(2\,\theta))\ x\mathbf{U} + \sin(2\,\theta)\,\mathbf{U}\times(x\mathbf{U} + y\mathbf{A} + z\mathbf{B})\\
=&\ \cos(2\,\theta)\,y\mathbf{A} + \cos(2\,\theta)\,z\mathbf{B} + x\mathbf{U} + \sin(2\,\theta)\,\mathbf{U}\times(y\mathbf{A}) + \sin(2\,\theta)\,\mathbf{U}\times(z\mathbf{B})\\
=&\ x\mathbf{U} + \cos(2\,\theta)\,y\mathbf{A} + \cos(2\,\theta)\,z\mathbf{B} + \sin(2\,\theta)\,y\mathbf{B} + \sin(2\,\theta)\ (-z\mathbf{A})\\
=&\ x\mathbf{U} + (\cos(2\,\theta)\,y - \sin(2\,\theta)\,z)\,\mathbf{A} + (\sin(2\,\theta)\,y + \cos(2\,\theta)\,z)\,\mathbf{B}
\end{aligned}$$

Expressed in terms of matrices, the vector $\vec{v}$ has coordinates (x, y, z) in basis $\mathbf{U}, \mathbf{A}, \mathbf{B}$. So the expression in the last line above is

$$\begin{pmatrix}1 & 0 & 0\\ 0 & \cos(2\,\theta) & -\sin(2\,\theta)\\ 0 & \sin(2\,\theta) & \cos(2\,\theta)\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}$$

in this basis, which makes it apparent that this represents rotation by angle $2\,\theta$ in the $\mathbf{AB}$ plane with axis $\mathbf{U}$, where $0 < 2\,\theta < 2\,\pi$.

If $\mathbf{P}$ is the matrix with columns $\mathbf{U}, \mathbf{A}, \mathbf{B}$ (in that order) then the matrix of this rotation is

$$\mathbf{M} = \mathbf{P}\begin{pmatrix}1 & 0 & 0\\ 0 & \cos(2\,\theta) & -\sin(2\,\theta)\\ 0 & \sin(2\,\theta) & \cos(2\,\theta)\end{pmatrix}\mathbf{P}^{-1}.$$

Of course this does not depend on choice of $\mathbf{A}$. Any other choice $\mathbf{K}$ perpendicular to $\mathbf{U}$ would generate a matrix of transition from the $\mathbf{U}, \mathbf{K}, \mathbf{U}\times\mathbf{K}$ basis to the $\mathbf{U}, \mathbf{A}, \mathbf{B}$ basis of the form

$$\begin{pmatrix}1 & 0 & 0\\ 0 & c & -s\\ 0 & s & c\end{pmatrix}\quad \text{where}\ \ c^2 + s^2 = 1\ \ \text{and we find that}$$

$$\begin{pmatrix}1 & 0 & 0\\ 0 & c & s\\ 0 & -s & c\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & \cos(2\,\theta) & -\sin(2\,\theta)\\ 0 & \sin(2\,\theta) & \cos(2\,\theta)\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & c & -s\\ 0 & s & c\end{pmatrix} = \begin{pmatrix}1 & 0 & 0\\ 0 & \cos(2\,\theta) & -\sin(2\,\theta)\\ 0 & \sin(2\,\theta) & \cos(2\,\theta)\end{pmatrix}$$

and the same orthogonal matrix $\mathbf{M}$ would be generated.

Note also that there are *two* different unit quaternions, identified with points on the unit sphere $\mathbb{S}^3$ in $\mathbb{R}^4$, which can be used to produce each rotation. Both q and $-q$ induce the same rotation, one corresponding to a counterclockwise rotation around an axis vector $\mathbf{U}$, the other a counterclockwise rotation around axis vector $-\mathbf{U}$. The sum of the two angles of rotation is $2\,\pi$. The "double cover" of the rotations by the unit quaternions is not just a mathematical quirk: it has physical significance, as will be pointed out in Section 23.

To reiterate, we know that each orthogonal matrix with determinant 1 is associated with exactly one rotation, which can be identified by an axis and rotation angle. On the other hand, each unit quaternion corresponds to a rotation and therefore exactly one rotation matrix. And knowledge of an axis of rotation and rotation angle can be used to generate two quaternions, each of which can be used to implement the rotation and, if you wish, create the associated orthogonal matrix.

Finally, we examine another way to produce the associated matrix from axis vector and angle by taking another look at

$$q\,\vec{v}\,q^* = \cos(2\,\theta)\,\vec{v} + (\,1 - \cos(2\,\theta))\ (\mathbf{U}\cdot\vec{v}\,)\ \mathbf{U} + \sin(2\,\theta)\,\mathbf{U}\times\vec{v}.$$

If generic vector $\vec{v}$ in $\mathbb{R}^3$ has coordinates (v_1, v_2, v_3) and axis vector $\mathbf{U}$ has coordinates (U_1, U_2, U_3), both written as columns, then a direct calculation shows that the above equation for the rotated vector $\mathbf{M}\,\vec{v}$ equals

$$\mathbf{M}\,\vec{v} = \left(\ \cos(2\,\theta)\,\mathbf{I} + (1-\cos(2\,\theta))\,\mathbf{U}\mathbf{U}^T + \sin(2\,\theta)\,\mathbf{U}^{\times}\ \right)\vec{v}$$

where $\mathbf{I}$ is the 3×3 identity matrix, $\mathbf{U}\mathbf{U}^T$ is a symmetric 3×3 matrix and $\mathbf{U}^{\times}$ is the 3×3 matrix that implements cross product with $\mathbf{U}$ (on the left):

$$\mathbf{U}\mathbf{U}^T = \begin{pmatrix} (U_1)^2 & U_1U_2 & U_1U_3 \\ U_2U_1 & (U_2)^2 & U_2U_3 \\ U_3U_1 & U_3U_2 & (U_3)^2 \end{pmatrix} \qquad \mathbf{U}^{\times} = \begin{pmatrix} 0 & -U_3 & U_2 \\ U_3 & 0 & -U_1 \\ -U_2 & U_1 & 0 \end{pmatrix}.$$

This provides a direct and somewhat faster way (without the choices involved in finding the matrix of transition $\mathbf{P}$) to create a matrix for the rotation around normalized axis $\mathbf{U}$ by angle $2\,\theta$. It is called **Rodrigues' formula**.

More importantly, if you actually have a rotation matrix $\mathbf{M}$ in hand, Rodrigues' formula shows that

$$\mathbf{M} - \mathbf{M}^T = 2\,\sin(2\,\theta)\,\mathbf{U}^{\times} = 2\,\sin(2\,\theta)\begin{pmatrix} 0 & -U_3 & U_2 \\ U_3 & 0 & -U_1 \\ -U_2 & U_1 & 0 \end{pmatrix}$$

so by normalizing you can determine the coordinates U_1, U_2 and U_3 of the unit rotation axis and the rotation angle $2\,\theta$.

12. Power Series and Derivatives for Quaternions

In the Section 10 we found that quaternions can be expressed as

$$q = a + b\,\mathbf{U}$$

where a and b are real and $\mathbf{U}$ is imaginary with $\mathbf{U}^2 = -1$.

If q is not real and we choose $\mathbf{U}$ so that $b > 0$ the representation is unique, but allowing negative coefficients gives the only two different representations of each q in terms of unit imaginary quaternions:

$$q = a + b\,\mathbf{U} = a + (-b)\ (-\mathbf{U})\,.$$

This means that if

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

is any real power series with radius of convergence R and if q is a quaternion and $\|q\| < R$ then the **power series** $f(q)$ can be defined by

$$f(q) = \sum_{n=0}^{\infty} a_n q^n$$

exactly in the way that f was extended to a complex argument. The problem here is that the lack of commutativity of $\mathbb{H}$ severely restricts what one can *do* with functions defined in this way.

For instance, suppose we define, for any quaternion q the **exponential function**

$$e^q = \sum_{n=0}^{\infty} \frac{q^n}{n!}.$$

If p is another quaternion and $p = c + d\,\mathbf{V}$ then unless the imaginary part of one is a multiple of the imaginary part of the other $p\,q \neq q\,p$ so we cannot say how $e^q e^p$ might be related to e^{q+p}.

It is true that *restricted to the slice* $\mathbb{H}_{\mathbf{U}}$ of $\mathbb{H}$ defined by unit imaginary quaternion $\mathbf{U}$ the function f *on that slice* satisfies all the usual properties. For instance $e^q e^p = e^{q+p}$ for quaternions p and q both in the slice $\mathbb{H}_{\mathbf{U}}$. Recall also that the real quaternions are in every slice (though that is the only overlap between unequal slices) so if $q = a + b\mathbf{U}$ for real a and b and imaginary $\mathbf{U}$ then

$$e^{a+b\mathbf{U}} = e^a e^{b\mathbf{U}} \quad \text{and} \quad \left(e^{a+b\mathbf{U}}\right)^* = e^{a-b\mathbf{U}}.$$

For any function f defined on quaternion q by a real power series, if q is in $\mathbb{H}_{\mathbf{U}}$ then so is $f(q)$. This of course is not the case for more general quaternion-valued functions.

Uniqueness of a real power series representation follows as in the complex case. Derivatives and integrals *restricted to this particular slice* exist, and can obtained by term-by-term operations.

For nonzero quaternion q with

$$q = \|q\|\,(\,\cos(\theta) + \sin(\theta)\,\mathbf{U}\,)$$

for unit imaginary $\mathbf{U}$ and real h and θ in $(h, h + 2\,\pi)$ we can define a branch of the log function on q by

$$\log(q) = \log(\|q\|) + \theta\,\mathbf{U}.$$

The angle θ can be denoted $\mathbf{arg(q)}$, called the **argument** of q. As in the complex case there is ambiguity here.

$\arg(q)$ depends on both h and $\mathbf{U}$, or rather h and choice of either $\mathbf{U}$ or $-\mathbf{U}$. This is because there is a second angle $\overline{\theta}$ in $(h, h + 2\,\pi)$

$$\cos(\theta) + \sin(\theta)\,\mathbf{U} = \cos\left(\overline{\theta}\right) + \sin\left(\overline{\theta}\right)(-\mathbf{U})$$

so there are potentially two arguments for those nonzero q in $\mathbb{H}_{\mathbf{U}}$ which do not lie along the ray at angle h. The two possibilities, θ and $\overline{\theta}$, add to a multiple of $2\,\pi$. We must specify which of these arguments we intend, and that means specifying $-\mathbf{U}$ or $\mathbf{U}$ as well as h. Each choice leads to a different branch of the logarithm.

Defining as above the exponential function by power series on the slice $\mathbb{H}_{\mathbf{U}}$ we have for $q = \|q\|\,(\,\cos(\theta) + \sin(\theta)\,\mathbf{U}\,)$

$$e^{\log(q)} = e^{\ln(\|q\|)+\theta\,\mathbf{U}} = e^{\log(\|q\|)} e^{\theta\,\mathbf{U}} = \|q\|\,e^{\theta\,\mathbf{U}} = \|q\|\,(\,\cos(\theta) + \sin(\theta)\,\mathbf{U}\,) = q.$$

Using the other branch of the logarithm defined with the same h but unit imaginary quaternion $-\mathbf{U}$ we have also

$$e^{\log(q)} = e^{\ln(\|q\|)+\overline{\theta}(-\mathbf{U})} = \|q\|\, e^{\overline{\theta}(-\mathbf{U})} = \|q\| \left(\cos\left(\overline{\theta}\right) - \sin\left(\overline{\theta}\right) \mathbf{U} \right) = q.$$

Finally, we will be particularly interested in *unit* quaternions, and we point out the obvious here: those can all be represented as

$$q = e^{\theta\,\mathbf{U}} = \cos(\theta) + \sin(\theta)\,\mathbf{U} \quad \text{ where } \mathbf{U}^2 = -1, \text{ and } \|q\| = 1.$$

With exponentials and logs in hand we can define other functions using them. For instance we can define q^p for q in the domain of a branch of the logarithm, by

$$\mathbf{q^p} = e^{\,p\log(q)} \qquad p, q \text{ both in some } \mathbb{H}_{\mathbf{U}}.$$

A further word about derivatives is in order. We mentioned that restricted to each slice quaternion functions are no different than any other representation of complex-valued functions of a complex variable.

Attempting to define a **derivative** *not* restricted to a slice, as is done for complex derivatives

$$f'(q) = \lim_{h\to 0} \frac{f(q+h) - f(q)}{h} \qquad (h \text{ in } \mathbb{H})$$

fails from the outset due to the fact that left division by h need not equal right division by h so the expression is ambiguous.

So let's agree to worry about that later and settle for now on right division by h and a function as simple as $f(q) = q^2$.

$$\begin{aligned}\left((q+h)^2 - q^2\right) h^{-1} = {} & \left(q^2 + q\,h + h\,q + h^2 - q^2\right) h^{-1} \\ & = q + h\,q\,h^{-1} + h = q + \frac{h}{\|h\|}\, q\, \frac{h^*}{\|h\|} + h.\end{aligned}$$

$h/\|h\|$ is a unit quaternion and in section 11 we found that the product

$$\frac{h}{\|h\|}\, q\, \frac{h^*}{\|h\|}$$

will rotate the imaginary part of q and the angle and axis of this rotation can be varied as the norm of h approaches 0.

In other words, the limit (and so the derivative) cannot exist unless q is real or h is confined to the same slice as q. The burden of satisfying the differentiability condition is so onerous that virtually no function can bear it.

So we will restrict our consideration of derivatives to functions f and quaternions q in $\mathbb{H}_{\mathbf{U}}$ for which $f(q+h)$ is in $\mathbb{H}_{\mathbf{U}}$ whenever h has sufficiently small norm and is itself in $\mathbb{H}_{\mathbf{U}}$. (Functions defined using real power series do satisfy this condition.) In this case h^{-1} will commute with $f(q+h) - f(q)$, so the left-right division choice is not an issue, and we define

$$\frac{d\,f}{d\,\mathbf{U}}(q) = \lim_{h\to 0} \frac{f(q+h) - f(q)}{h} \qquad (h \text{ in } \mathbb{H}_{\mathbf{U}})$$

whenever this limit exists.

As mentioned earlier, the existence of this limit is exactly analogous to the existence of a complex derivative (since $\mathbb{H}_{\mathbf{U}}$ is algebra isomorphic to $\mathbb{C}$) and the existence of a complex derivative is a very strong condition. All the usual powerful theorems about complex differentiable functions apply on each slice.

It is possible to create in some cases the idea of **directional derivative** for quaternion valued functions of a quaternion variable, a weaker notion of differentiability but sufficient to give us some useful information about, for instance, the location of potential minima or other analogues from ordinary calculus.

The **Gâteaux derivative** of f at q in the direction of nonzero p is defined as

$$D_p f(q) = \lim_{t\to 0} \frac{f(q+tp) - f(q)}{t} \qquad (t \text{ real})$$

when the limit exists.

If q and nonzero p are both in $\mathbb{H}_{\mathbf{U}}$ and if $\frac{df}{d\,\mathbf{U}}$ does exist at q, the Gâteaux derivative $D_p(q)$ will exist and

$$\frac{df}{d\,\mathbf{U}}(q) = \frac{D_p f(q)}{p}.$$

Finally, the **derivative of a path in the quaternions** (a quaternion valued function given for real numbers in an interval) is defined as would be done for any function into $\mathbb{R}^4$.

Specifically, if f is defined on the interval (a,b) and has value

$$f(c) = f_0(c) + f_1(c)\,i + f_2(c)\,j + f_3(c)\,k$$

for each c is in the interval we define $\frac{\mathbf{df}}{\mathbf{dt}}(\mathbf{c})$ to be $f_0'(c) + f_1'(c)\,i + f_2'(c)\,j + f_3'(c)\,k$ whenever all four ordinary component derivatives exist.

Both this derivative and the Gâteaux derivative are real linear, and satisfy the product rule with Hamilton product when derivatives of each factor exist.

13. Quaternions as Matrices

Consider the four matrices:

$$\mathbf{E} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The reader should check that

$$\mathbf{A}^2 = -\mathbf{E} \quad \text{and} \quad \mathbf{B}^2 = \mathbf{C}^2 = \mathbf{E}$$

$$\text{while} \quad \mathbf{AB} = -\mathbf{BA} = \mathbf{C} \quad \text{and} \quad \mathbf{CB} = -\mathbf{BC} = \mathbf{A} \quad \text{and} \quad \mathbf{CA} = -\mathbf{AC} = \mathbf{B}.$$

Using this "multiplication table" you can avoid matrix multiplications as you work with these examples.

Recall that we saw that the collection of matrices of the form

$$x\mathbf{E} + y\mathbf{A} \qquad \text{for real } x \text{ and } y$$

is algebra isomorphic to (i.e., "is," or "is a representation of") the complex numbers.

Commonly, one sees the quaternions represented as the set of all matrices which can be written as a combination, with real w, x, y and z,

$$\begin{aligned}&w\,\mathbf{E} + x\,(-i\mathbf{B}) + y\,(-\mathbf{A}) + z\,(-i\mathbf{C})\\ &= w\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + x\begin{pmatrix}-i & 0\\ 0 & i\end{pmatrix} + y\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} + z\begin{pmatrix}0 & -i\\ -i & 0\end{pmatrix}\\ &= \begin{pmatrix}w & 0\\ 0 & w\end{pmatrix} + \begin{pmatrix}-ix & y-iz\\ -y-iz & ix\end{pmatrix} = \begin{pmatrix}w-ix & y-iz\\ -y-iz & w+ix\end{pmatrix}.\end{aligned}$$

The order $-i\mathbf{B} \Rightarrow -\mathbf{A} \Rightarrow -i\mathbf{C}$ is not required here. A variety of permutations with appropriate sign changes would serve just as well.

If you multiply two sums of this kind together, you get another sum of this kind, and the same is true, more obviously, if you add two such sums, or multiply a sum like this by a scalar. So these matrices constitute a 4 dimensional real algebra. Associativity is automatic: matrix multiplication is associative.

Satisfy yourself that the function that sends a generic quaternion $w+x\,i+y\,j+z\,k$ to the matrix shown above is an algebra isomorphism. The squared norm of the quaternion is the determinant of this matrix, just as with the matrix representation of complex numbers. The conjugate of the quaternion is the conjugate transpose of the matrix of that quaternion.

If you want to represent the quaternions in terms of *real* matrices only, simply substitute in these matrices the block matrices

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \text{ in place of } 1, \text{ and } \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix} \text{ in place of } 0, \text{ and } \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} \text{ in place of } i.$$

This gives one representation of quaternion $q = w + x\,i + y\,j + z\,k$ as

$$w + x\,i + y\,j + z\,k \Longleftrightarrow \begin{pmatrix}w & x & y & z\\ -x & w & -z & y\\ & & & \\ -y & z & w & -x\\ -z & -y & x & w\end{pmatrix}.$$

However, perhaps the nicest representation of quaternions in terms of real matrices is in terms of the 2×2 block matrices

$$w\begin{pmatrix}\mathbf{E} & \mathbf{0}\\ \mathbf{0} & \mathbf{E}\end{pmatrix} + x\begin{pmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{A}\end{pmatrix} + y\begin{pmatrix}\mathbf{0} & -\mathbf{B}\\ \mathbf{B} & \mathbf{0}\end{pmatrix} + z\begin{pmatrix}\mathbf{0} & -\mathbf{C}\\ \mathbf{C} & \mathbf{0}\end{pmatrix} = \begin{pmatrix}w & -x & -y & -z\\ x & w & -z & y\\ & & & \\ y & z & w & -x\\ z & -y & x & w\end{pmatrix}$$

The conjugate quaternion q^* corresponds to the transpose of this matrix, and a calculation shows the determinant is

$$\left(w^2 + x^2 + y^2 + z^2\right)^2 = \|q\|^4.$$

If you identify generic quaternion $a + b\,i + c\,j + d\,k$ with the member (a, b, c, d) of $\mathbb{R}^4$ written as a column, then left multiplication by quaternion $w + x\,i + y\,j + z\,k$

corresponds to a linear transformation on $\mathbb{R}^4$ and hence can be represented by left multiplication by a matrix. The matrix we just presented is the one.

$$(w + x\,i + y\,j + z\,k)\;(a + b\,i + c\,j + d\,k)$$

$$\Longleftrightarrow \begin{pmatrix} aw - xb - yc - zd \\ xa + wb - zc + yd \\ ya + zb + wc - xd \\ za - yb + xc + wd \end{pmatrix} = \begin{pmatrix} w & -x & -y & -z \\ x & w & -z & y \\ y & z & w & -x \\ z & -y & x & w \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

$$\Longleftrightarrow \begin{pmatrix} w & -x & -y & -z \\ x & w & -z & y \\ y & z & w & -x \\ z & -y & x & w \end{pmatrix} \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix}.$$

Of course right multiplication is also linear, but written on the left as matrix product we have the slightly different:

$$(w + x\,i + y\,j + z\,k)\;(a + b\,i + c\,j + d\,k) \Longleftrightarrow \begin{pmatrix} a & -b & -c & -d \\ b & a & d & -c \\ c & -d & a & b \\ d & c & -b & a \end{pmatrix} \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}.$$

14. The Pauli Spin Matrices

In a bit of a side note to our main purpose, we define the matrices

$$\sigma_0 = \mathbf{E} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \sigma_1 = \mathbf{C} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\text{and} \quad \sigma_2 = i\mathbf{A} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{and} \quad \sigma_3 = \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

so that

$$\sigma_1\sigma_2 = i\sigma_3 \quad \text{and} \quad \sigma_2\sigma_3 = i\sigma_1 \quad \text{and} \quad \sigma_3\sigma_1 = i\sigma_2.$$

and also

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_0 = e \quad \text{with} \quad \sigma_k\sigma_m = -\sigma_m\sigma_k \ \text{ for nonzero unequal } m \text{ and } k.$$

The matrices σ_1, σ_2 and σ_3 are called **Pauli spin matrices**. They are hermitian, which means they are their own conjugate transpose, and so correspond to "observables" in quantum mechanics. They represent spin with respect to the coordinate axes in quantum mechanical descriptions of certain particles. Any 2 by 2 hermitian matrix can be found (in one way) as a combination

$$w\,\sigma_0 + x\,\sigma_1 + y\,\sigma_2 + z\,\sigma_3 \;=\; \begin{pmatrix} w + z & x - iy \\ x + iy & w - z \end{pmatrix}$$

with real w, x, y and z.

These matrices constitute a representation of one of the smaller Clifford (also called geometric) algebras, a structure which we will resist exploring in these notes.

One easily shows that

$$\begin{aligned} w\,\sigma_0 + x\,(-i\sigma_1) + y\,(-i\sigma_2) &+ z\,(-i\sigma_3) \\ &= w\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + x\begin{pmatrix}0 & -i\\ -i & 0\end{pmatrix} + y\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} + z\begin{pmatrix}-i & 0\\ 0 & i\end{pmatrix} \\ &= \begin{pmatrix} w - zi & -xi - y \\ -xi + y & w + zi\end{pmatrix} \end{aligned}$$

with real w, x, y and z provides yet another way of realizing the quaternions, this time in terms of the Pauli spin matrices.

Here too the conjugate transpose matrix corresponds to the conjugate quaternion, and the norm squared is the matrix determinant.

As in the last section, writing in terms of 4 by 4 real matrices, the representation of quaternion $q = w + x\,i + y\,j + z\,k$ is

$$w\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix} + x\begin{pmatrix}0 & 0 & 0 & 1\\ 0 & 0 & -1 & 0\\ 0 & 1 & 0 & 0\\ -1 & 0 & 0 & 0\end{pmatrix}$$

$$+\,y\begin{pmatrix}0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{pmatrix} + z\begin{pmatrix}0 & 1 & 0 & 0\\ -1 & 0 & 0 & 0\\ 0 & 0 & 0 & -1\\ 0 & 0 & 1 & 0\end{pmatrix}$$

yielding

$$\begin{pmatrix}w & z & -y & x\\ -z & w & -x & -y\\ y & x & w & -z\\ -x & y & z & w\end{pmatrix} = \begin{pmatrix}w & 0 & 0 & 0\\ 0 & w & 0 & 0\\ 0 & 0 & w & 0\\ 0 & 0 & 0 & w\end{pmatrix} + \begin{pmatrix}0 & z & -y & x\\ -z & 0 & -x & -y\\ y & x & 0 & -z\\ -x & y & z & 0\end{pmatrix}.$$

The conjugate operation corresponds to transpose, and a calculation shows the determinant is the fourth power of the norm as before.

15. Rotations in $\mathbb{R}^4$

A linear rigid motion $\boldsymbol{\mathcal{L}}$ on $\mathbb{R}^4$ is a linear map that sends any (and hence every) right-handed ordered orthonormal basis to another. It sends the **standard basis**

$$\vec{\boldsymbol{e}}_1 = \langle 1,0,0,0\rangle\,, \quad \vec{\boldsymbol{e}}_2 = \langle 0,1,0,0\rangle\,, \quad \vec{\boldsymbol{e}}_3 = \langle 0,0,1,0\rangle\,, \quad \vec{\boldsymbol{e}}_4 = \langle 0,0,0,1\rangle\,,$$

to an orthonormal basis $\vec{u}_1$, $\vec{u}_2$, $\vec{u}_3$, $\vec{u}_4$ and the matrix $\mathbf{M} = (\vec{u}_1\,\vec{u}_2\,\vec{u}_3\,\vec{u}_4)$, which is the matrix of this linear transformation, must have determinant 1.

$$\boldsymbol{\mathcal{L}}(v) = \mathbf{M}v \quad \text{for every vector } v \text{ in } \mathbb{R}^4.$$

Since $\mathbf{M}$ is real any complex roots come in conjugate pairs. Since $\mathbf{M}$ is an isometry, the norm of each root is 1.

If the characteristic polynomial of $\mathbf{M}$ has four real roots, it could have all four equal to 1, exactly two equal to -1 or all four -1. It is also possible that there could be two real roots and a conjugate pair of complex roots, or there could be two sets of conjugate pairs of roots.

Because $\mathbf{M}$ is normal there is a right-handed ordered orthonormal basis

$$\vec{p}_1,\, \vec{p}_2,\, \vec{p}_3, \vec{p}_4$$

so that the orthogonal matrix of transition $\mathbf{P}$, whose columns are this basis, satisfies

$$\mathbf{D}_{\mu,\theta} = \mathbf{P}^{-1}\mathbf{M}\mathbf{P} = \begin{pmatrix} \cos(\mu) & -\sin(\mu) & 0 & 0 \\ \sin(\mu) & \cos(\mu) & 0 & 0 \\ 0 & 0 & \cos(\theta) & -\sin(\theta) \\ 0 & 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}$$

for certain θ and μ in $(-\pi, \pi]$.

In the first case, root 1 repeated four times, $\mathbf{M}$ is the identity matrix and $\mu = \theta = 0$. In the second case we can choose $\mu = 0$ and $\theta = \pi$. In the third case both μ and θ are π and $\mathbf{M}$ represents inversion. Unlike 3 dimensions, but similar to 2 dimensions, inversion is rigid in $\mathbb{R}^4$.

If there is a single pair of complex roots, the angle μ could be 0 or π in which case $\mathbf{M}$ acts as the identity or inversion in the plane of $\vec{p}_1$ and $\vec{p}_2$, and rotation by angle θ in the plane of $\vec{p}_3$ and $\vec{p}_4$.

In the final case, $\mathbf{M}$ acts as rotation by angle μ in the plane of $\vec{p}_1$ and $\vec{p}_2$, and by angle θ in the plane of $\vec{p}_3$ and $\vec{p}_4$.

16. Quaternions and Isometries on $\mathbb{R}^4$

Suppose α and β are quaternions, conflated with members of $\mathbb{R}^4$ as

$$\alpha = w + x\,i + y\,j + z\,k \;\Leftrightarrow\; \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}$$

whenever that is convenient.

We define the linear function F on β in $\mathbb{H}$ by

$$F_\alpha(\beta) = -\alpha\,\overline{\beta}\,\alpha \quad \text{where } \alpha \text{ is a fixed unit quaternion.}$$

The quaternions α, $i\,\alpha$, $j\,\alpha$ and $k\,\alpha$ form an orthonormal basis of $\mathbb{H}$, so the last three vectors span the three dimensional subspace $\alpha^\perp$ of $\mathbb{H}$.

A calculation verifies that

$$F_\alpha(\alpha) = -\alpha, \quad F_\alpha(i\,\alpha) = i\,\alpha, \quad F_\alpha(j\,\alpha) = j\,\alpha \;\text{ and }\; F_\alpha(k\,\alpha) = k\,\alpha.$$

So F_α is a reflection in the hyperplane perpendicular to α.

We discovered in Section 3 that any linear isometry in $\mathbb{R}^4$ can be realized as the composition of no more than four reflections, and nontrivial linear rigid motions (isometries with determinant 1) are the composition of either two or four reflections.

That means linear rigid motions applied to β are of the form

$$(\alpha_2\,\overline{\alpha_1})\;\beta\;(\overline{\alpha_1}\,\alpha_2)\quad\text{or}\quad(\alpha_4\,\overline{\alpha_3}\,\alpha_2\,\overline{\alpha_1})\;\beta\;(\overline{\alpha_1}\,\alpha_2\,\overline{\alpha_3}\,\alpha_4)$$

while other nontrivial linear isometries have the form

$$-\alpha_1\,\overline{\beta}\,\alpha_1\quad\text{or}\quad-\,(\alpha_3\,\overline{\alpha_2}\,\alpha_1)\;\overline{\beta}\;(\alpha_1\,\overline{\alpha_2}\,\alpha_3)$$

for various unit quaternions α_i.

For us, the important point to take away from this calculation is:

Every linear rigid motion of $\mathbb{R}^4$ can be calculated on quaternion β, thought of as a member of $\mathbb{R}^4$, as the Hamilton product $\alpha\,\beta\,\gamma$ for certain unit quaternions α and γ.

Every orthogonal matrix with determinant 1 can be construed as a linear rigid motion, so left multiplication by such a matrix can be calculated using quaternions in this way.

Representing quaternions as columns in $\mathbb{R}^4$, we saw at the end of Section 13 that if three generic quaternions (not necessarily unit quaternions)

$$\alpha \Leftrightarrow \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}, \quad \beta \Leftrightarrow \begin{pmatrix} l \\ m \\ n \\ o \end{pmatrix}, \quad \gamma \Leftrightarrow \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

are multiplied, then

$$(\alpha\beta)\gamma \Leftrightarrow \begin{pmatrix} a & -b & -c & -d \\ b & a & d & -c \\ c & -d & a & b \\ d & c & -b & a \end{pmatrix} \begin{pmatrix} w & -x & -y & -z \\ x & w & -z & y \\ y & z & w & -x \\ z & -y & x & w \end{pmatrix} \begin{pmatrix} l \\ m \\ n \\ o \end{pmatrix}.$$

Associated in the other order we have

$$\alpha(\beta\gamma) \Leftrightarrow \begin{pmatrix} w & -x & -y & -z \\ x & w & -z & y \\ y & z & w & -x \\ z & -y & x & w \end{pmatrix} \begin{pmatrix} a & -b & -c & -d \\ b & a & d & -c \\ c & -d & a & b \\ d & c & -b & a \end{pmatrix} \begin{pmatrix} l \\ m \\ n \\ o \end{pmatrix}$$

so matrix pairs of this particular form commute.

Let $\boldsymbol{\mathcal{L}eft}$ denote the set of matrices which can be produced by left multiplication by a quaternion, as above, and $\boldsymbol{\mathcal{R}ight}$ denote those matrices which can be formed to represent right multiplication.

Inspection shows that the only matrices in both $\boldsymbol{\mathcal{L}eft}$ and $\boldsymbol{\mathcal{R}ight}$ are scalar multiples of the identity matrix.

If the matrices corresponding to left and right multiplication by quaternion α are denoted L_α and R_α, respectively, and if for all β

$$\alpha_1\beta\gamma_1 = \alpha_2\beta\gamma_2$$

then $\alpha\beta\gamma = \beta$ for all β where $\alpha = \alpha_2^{-1}\alpha_1$ and $\gamma = \gamma_1\gamma_2^{-1}$. It follows that the product $L_\alpha R_\gamma$ is the identity matrix. But then $L_\alpha^{-1} = L_{\alpha^{-1}} = R_\gamma$ and this implies that L_α and R_γ are in both $\boldsymbol{\mathcal{R}ight}$ and $\boldsymbol{\mathcal{L}eft}$: in other words, they are multiples of the identity matrix.

Our conclusion: there is a nonzero real constant k so that $\alpha_1 = k\,\alpha_2$ and $\gamma_1 = k^{-1}\,\gamma_2$. So any (nonzero) matrix that *can* be represented as a product $L_\alpha R_\gamma$ can be so represented as $L_{(k\,\alpha)}R_{(k^{-1}\,\gamma)}$ for nonzero real k and in no other way.

Some observations: (i) The determinants of the matrices L_α and R_α are both $\|\alpha\|^4$. (ii) Unless $\alpha = 0$ the four columns of each matrix form orthogonal sets of vectors.

So if α and γ are unit quaternions, the matrices L_α and R_γ and $L_\alpha R_\gamma = R_\gamma L_\alpha$ are matrices of rotations of $\mathbb{R}^4$ and the two quaternion products $\alpha\beta\gamma$ and $(-\alpha)\beta(-\gamma)$ can be thought of as implementing this product rotation on β as a member of $\mathbb{R}^4$. There are no other unit quaternion pairs that could.

And, as we saw above, any orthogonal matrix with determinant 1 *can* be represented in this way.

We will now examine some consequences of this fact.

Suppose $\mathbf{P}$ is any orthogonal matrix with determinant 1 and $\mathbf{L} \in \boldsymbol{\mathcal{L}eft}$. Then $\mathbf{P} = L_\alpha R_\gamma$ for certain α and γ so

$$\mathbf{P}^{-1}\,\mathbf{L}\,\mathbf{P} = R_{\gamma^{-1}}\,L_{\alpha^{-1}}\,\mathbf{L}\,L_\alpha\,R_\gamma = L_{\alpha^{-1}}\,R_{\gamma^{-1}}\,R_\gamma\,\mathbf{L}\,L_\alpha = L_{\alpha^{-1}}\,\mathbf{L}\,L_\alpha$$

and since $\boldsymbol{\mathcal{L}eft}$ is closed under multiplication this last is itself a member of $\boldsymbol{\mathcal{L}eft}$. The same is true for $\boldsymbol{\mathcal{R}ight}$: both $\boldsymbol{\mathcal{L}eft}$ and $\boldsymbol{\mathcal{R}ight}$ are closed under change of basis to any other orthonormal right-handed basis.

Now suppose $\mathbf{M} = \mathbf{LR}$ with $\mathbf{L} \in \boldsymbol{\mathcal{L}eft}$ and $\mathbf{R} \in \boldsymbol{\mathcal{R}ight}$.

Then we have

$$\begin{aligned}\mathbf{P}^{-1}\,\mathbf{M}\,\mathbf{P} &= \mathbf{P}^{-1}\,\mathbf{L}\,\mathbf{R}\,\mathbf{P} = \mathbf{P}^{-1}\,\mathbf{L}\,\mathbf{P}\,\mathbf{P}^{-1}\,\mathbf{R}\,\mathbf{P} \\ &= R_{\gamma^{-1}}\,L_{\alpha^{-1}}\,\mathbf{L}\,L_\alpha\,R_\gamma\,R_{\gamma^{-1}}\,L_{\alpha^{-1}}\,\mathbf{R}\,L_\alpha\,R_\gamma = \left(L_{\alpha^{-1}}\,\mathbf{L}\,L_\alpha\right)\left(R_{\gamma^{-1}}\,\mathbf{R}\,R_\gamma\right).\end{aligned}$$

The second line is interesting, but the point of this lies in the first.

Every orthogonal matrix $\mathbf{M}$ with determinant 1 can be reduced to form $\mathbf{D}_{\mu,\theta}$ as defined in Section 15 using orthogonal matrix of transition $\mathbf{P}$, as $\mathbf{P}^{-1}\,\mathbf{M}\,\mathbf{P}$. The first line shows how to find the quaternions for $\mathbf{L}$ and $\mathbf{R}$ *if we can find the quaternions that produce* $\mathbf{D}_{\mu,\theta}$.

The matrices that corresponds to left multiplication by quaternion $\cos(\rho) + \sin(\rho)\,i = C + S\,i$ and right multiplication by quaternion $\cos(\tau) + \sin(\tau)\,i = c + s\,i$ are, respectively,

$$L(\rho) = \begin{pmatrix} C & -S & 0 & 0 \\ S & C & 0 & 0 \\ 0 & 0 & C & -S \\ 0 & 0 & S & C \end{pmatrix} \quad \text{and} \quad R(\tau) = \begin{pmatrix} c & -s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & c & s \\ 0 & 0 & -s & c \end{pmatrix}.$$

Their product is

$$\begin{pmatrix} C & -S & 0 & 0 \\ S & C & 0 & 0 \\ 0 & 0 & C & -S \\ 0 & 0 & S & C \end{pmatrix} \begin{pmatrix} c & -s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & c & s \\ 0 & 0 & -s & c \end{pmatrix}$$

$$= \begin{pmatrix} Cc - Ss & -Cs - Sc & 0 & 0 \\ Sc + Cs & Cc - Ss & 0 & 0 \\ 0 & 0 & Cc + Ss & Cs - Sc \\ 0 & 0 & Sc - Cs & Ck + Ss \end{pmatrix}$$

$$= \begin{pmatrix} \cos(\rho+\tau) & -\sin(\rho+\tau) & 0 & 0 \\ \sin(\rho+\tau) & \cos(\rho+\tau) & 0 & 0 \\ 0 & 0 & \cos(\rho-\tau) & -\sin(\rho-\tau) \\ 0 & 0 & \sin(\rho-\tau) & \cos(\rho-\tau) \end{pmatrix}$$

One can make this equal to

$$\mathbf{D}_{\mu,\theta} = \mathbf{P}^{-1}\mathbf{M}\mathbf{P} = \begin{pmatrix} \cos(\mu) & -\sin(\mu) & 0 & 0 \\ \sin(\mu) & \cos(\mu) & 0 & 0 \\ 0 & 0 & \cos(\theta) & -\sin(\theta) \\ 0 & 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}$$

by solving the equations

$$\rho + \tau = \mu \quad \text{and} \quad \rho - \tau = \theta.$$

Then

$$M = \mathbf{P}\,\mathbf{D}_{\mu,\theta}\,\mathbf{P}^{-1} = \mathbf{P}\,L(\rho)\,R(\tau)\,\mathbf{P}^{-1} = \left(\mathbf{P}\,L(\rho)\,\mathbf{P}^{-1}\right)\,\left(\mathbf{P}\,R(\tau)\,\mathbf{P}^{-1}\right)$$

and we read off the two quaternions we are looking for as the leftmost columns of

$$\mathbf{L} = \mathbf{P}\,L(\rho)\,\mathbf{P}^{-1} \quad \text{and} \quad \mathbf{R} = \mathbf{P}\,R(\tau)\,\mathbf{P}^{-1}.$$

Recall from above that the change of basis can actually be implemented as

$$\mathbf{L} = L_\alpha\,L(\rho)\,L_{\alpha^{-1}} \quad \text{and} \quad \mathbf{R} = R_\gamma\,R(\tau)\,R_{\gamma^{-1}}$$

for certain unit quaternions $\alpha = e^{x\,\mathbf{U}}$ and $\gamma = e^{y\,\mathbf{V}}$ at angles x and y and unit imaginaries $\mathbf{U}$ and $\mathbf{V}$.

So if β is any quaternion with associated vector B in $\mathbb{R}^4$ as

$$\beta = l + m\,i + n\,j + o\,k \quad \Longleftrightarrow \quad B = \begin{pmatrix} l \\ m \\ n \\ 0 \end{pmatrix}$$

then we can rotate B, finding the rotated vector $\mathbf{M}B$, by doing the quaternion calculation

$$\left(e^{x\,\mathbf{U}}\,e^{\rho\,i}\,e^{-x\,\mathbf{U}}\right)\,\beta\,\left(e^{-y\,\mathbf{V}}\,e^{\tau\,i}\,e^{y\,\mathbf{V}}\right).$$

It may be obvious that the left factor remains at angle ρ (for a different unit imaginary quaternion) and the right factor remains at angle τ, but we would like to confirm this.

Suppose $\mathbf{W}$ is any unit imaginary quaternion. (In the case we care about $\mathbf{W} = i$.) To ease the clutter we let $a = \cos(x)$, $b = \sin(x)$, $c = \cos(\rho)$ and $s = \sin(\rho)$.

$$\begin{aligned}
&(a + b\,\mathbf{U})\ (c + s\,\mathbf{W})\ (a - b\,\mathbf{U}) \\
&\qquad = (a + b\,\mathbf{U})\,[\,(ac + bs\,\mathbf{W}\cdot\mathbf{U}) + (-bc\,\mathbf{U} + as\,\mathbf{W} - bs\,\mathbf{W}\times\mathbf{U})\,] \\
&\qquad = a^2c + abs\,\mathbf{W}\cdot\mathbf{U} + b^2c\,\mathbf{U}\cdot\mathbf{U} - abs\,\mathbf{U}\cdot\mathbf{W} + b^2s\,\mathbf{U}\cdot(\mathbf{W}\times\mathbf{U}) \\
&\qquad\qquad + abc\,\mathbf{U} + b^2s\ (\mathbf{W}\cdot\mathbf{U})\,\mathbf{U} - abc\,\mathbf{U} + a^2s\,\mathbf{W} - abs\,\mathbf{W}\times\mathbf{U} \\
&\qquad\qquad\quad - b^2c\,\mathbf{U}\times\mathbf{U} + abs\,\mathbf{U}\times\mathbf{W} - b^2s\,\mathbf{U}\times(\mathbf{W}\times\mathbf{U}) \\
&\qquad = c + s\left[\ b^2\ (\mathbf{W}\cdot\mathbf{U})\,\mathbf{U} + a^2\,\mathbf{W} + 2ab\,\mathbf{U}\times\mathbf{W} + b^2\,\mathbf{U}\times(\mathbf{U}\times\mathbf{W})\ \right] \\
&\qquad = \ \cos(\rho) + \sin(\rho)\left[\ 2b^2\ (\mathbf{W}\cdot\mathbf{U})\,\mathbf{U} + \left(a^2 - b^2\right)\ \mathbf{W} + 2ab\,\mathbf{U}\times\mathbf{W}\ \right]
\end{aligned}$$

where the last line uses the triple cross product identity. The imaginary part on the right in that line *must* be a unit vector, and it is seen to be so by finding its norm, remembering that

$$(\mathbf{W}\cdot\mathbf{U})^2 + \|\mathbf{W}\times\mathbf{U}\|^2 = 1 \quad \text{and} \quad a^2 + b^2 = 1.$$

So we have verified the semi-obvious: the angles of the two rotation quaternions are the angles we calculated and used in $L(\rho)$ and $R(\tau)$ to create the block-diagonal $\mathbf{D}_{\mu,\theta}$, with $\rho + \tau = \mu$ and $\rho - \tau = \theta$. And there are unit imaginary quaternions $\mathbf{A}$ and $\mathbf{B}$ so that

$$e^{\rho\,\mathbf{A}}\,\beta\,e^{\tau\,\mathbf{B}}$$

implements the rotation on the quaternion β.

17. (*) Another Look at Rotations in $\mathbb{R}^3$

We have assembled a variety of representations for *unit* quaternion q, such as

$$q = w + x\,i + y\,j + z\,k = \cos(\theta) + \sin(\theta)\,\mathbf{U} = e^{\theta\,\mathbf{U}},$$

where $\mathbf{U}$ is a unit imaginary and θ is the corresponding argument.

Let's examine the matrix for a rotation in $\mathbb{R}^4$ corresponding to the linear function, created from unit quaternion q, given by the formula

$$\boldsymbol{\mathcal{F}}(p) = q\,p\,q^* = e^{\theta\,\mathbf{U}}\,p\,e^{-\theta\,\mathbf{U}} \quad \text{for every quaternion } p \text{ in } \mathbb{H}.$$

This is the situation that, if p is restricted to the imaginary quaternions, corresponds to a rotation in $\mathbb{R}^3$ by angle $2\,\theta$ around axis given by $\mathbf{U}$.

In $\mathbb{R}^4$ the corresponding function for unit q is

$$\boldsymbol{\mathcal{L}}(v) = \mathbf{M}\,v = L_q\,R_{q^*}\,v \quad \text{for every vector } v \text{ in } \mathbb{R}^4.$$

$\mathbf{M}$ is the product matrix $L_q\, R_{q^*}$ which is

$$L_q R_{q^*} = \begin{pmatrix} w & -x & -y & -z \\ x & w & -z & y \\ y & z & w & -x \\ z & -y & x & w \end{pmatrix} \begin{pmatrix} w & x & y & z \\ -x & w & z & -y \\ -y & -z & w & x \\ -z & y & -x & w \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-2y^2-2z^2 & 2xy-2wz & 2xz+2wy \\ 0 & 2xy+2wz & 1-2x^2-2z^2 & 2yz-2wx \\ 0 & 2xz-2wy & 2yz+2wx & 1-2x^2-2y^2 \end{pmatrix}.$$

The block matrix

$$\begin{pmatrix} 1-2y^2-2z^2 & 2xy-2wz & 2xz+2wy \\ 2xy+2wz & 1-2x^2-2z^2 & 2yz-2wx \\ 2xz-2wy & 2yz+2wx & 1-2x^2-2y^2 \end{pmatrix}$$

is the rotation matrix in $\mathbb{R}^3$ which corresponds to this rotation restricted to the imaginary quaternions. It can be calculated from the 9 products

$$x^2,\ y^2,\ z^2,\ wx,\ wy,\ wz,\ xy,\ xz \text{ and } yz$$

and is an alternative to the formula for this matrix from Section 11.

18. (*) Recap: Isometries in $\mathbb{R}^n$, for $n = 1, 2, 3$ and 4

There are two isometries that fix the origin in $\mathbb{R}$. The first is the identity, the second is the function $f(x) = -x$. The units of $\mathbb{R}$ consist of just the zero-dimensional "unit sphere" $\mathbb{S}^0 = \{1, -1\}$, and there is a unique isometry defined by left multiplication by each of these two units.

$\mathbb{R}^2$ can be identified with (or *is*) a normed algebra $\mathbb{C}$. The units of $\mathbb{C}$ correspond to the unit circle $\mathbb{S}^1$ in $\mathbb{R}^2$. Each origin-fixing isometry of $\mathbb{R}^2$ can be identified with the function given by $f_c(x) = cx$ for a unique c in $\mathbb{S}^1$.

Let's consider the unit sphere $\mathbb{S}^3$ in $\mathbb{R}^4$ and identify $\mathbb{R}^4$ with $\mathbb{H}$ and also identify $\mathbb{R}^3$ with the subspace of imaginary quaternions in $\mathbb{H}$.

If T is any origin-fixing isometry on $\mathbb{R}^3$ and if the determinant of T is 1 then there is a member q of $\mathbb{S}^3$ which implements the isometry via

$$T(x) = q\, x\, q^*.$$

If q produces the isometry T in this way, so does $-q$ but that is the only duplication.

On the other hand, if T is any linear isometry on $\mathbb{R}^3$ with determinant -1 then there is a member q of $\mathbb{S}^3$ which implements the isometry via

$$T(x) = q\, x^*\, q^*.$$

As before, if q produces the isometry T in this way, so does $-q$ but that is the only duplication.

Finally, we consider the case of $\mathbb{R}^4$.

If T is any linear isometry on $\mathbb{R}^4$ then there are members q and p of $\mathbb{S}^3$ so that

$$T(x) = q\, x\, p$$

if determinant of T is 1, and

$$T(x) = q\,x^*\,p$$

if determinant of T is -1.

If q and p produce the isometry T in this way, so does the pair $-q$ and $-p$, but this is the only duplication.

19. Octonions

We will take a brief detour to review a few properties of the remaining finite dimensional normed division algebra, denoted by $\mathbb{O}$ and usually called now the **octonions**. They were invented by **John Graves** in late December, 1843, following an exchange of letters with his friend Hamilton in which Hamilton outlined his newly found quaternions.

It is an eight dimensional real algebra, and can be defined using a procedure called the **Cayley-Dickson construction**, which can also be used to define $\mathbb{C}$ and $\mathbb{H}$, as follows.

We start with $\mathbb{R}$ and define a product on $\mathbb{R} \times \mathbb{R}$ by

$$(a, b)(x, y) = (ax - yb,\, ya + bx) \quad \text{for pairs of real numbers } \ (a, b) \ \text{ and } \ (x, y).$$

This is, of course, nothing more than complex multiplication. The conjugation operation applied to (a, b) is $(a, -b)$ and the complex norm corresponds to the normal Euclidean norm on $\mathbb{R} \times \mathbb{R}$.

We take the next step and define a product on $\mathbb{C} \times \mathbb{C}$ by

$$(a, b)(x, y) = (ax - \overline{y}b,\, ya + b\overline{x})$$

for pairs of complex numbers (a, b) and (x, y).

A calculation shows that the complex number pairs $(1, 0)$, $(i, 0)$, $(0, 1)$ and $(0, i)$ span this four dimensional real vector space and can be identified with the unit quaternions 1, i, j and k, respectively, to produce an algebra isomorphism onto the quaternions.

When $q = q_0(1, 0) + q_1\,(i, 0) + q_2\,(0, 1) + q_3\,(0, i)$ for real q_0, q_1, q_2 and q_3, the conjugate quaternion corresponds to $q^* = q_0(1, 0) - q_1\,(i, 0) - q_2\,(0, 1) - q_3\,(0, i)$ and the quaternion norm is then, of course, the square root of $q\,q^*$ and the multiplicative inverse of nonzero q is $q^*/\,(q\,q^*)$.

The octonions structure is produced by taking this procedure one more step.

Define a product on $\mathbb{O} = \mathbb{H} \times \mathbb{H}$ by

$$(a, b)(x, y) = (ax - y^*b,\, ya + b\,x^*)$$

for two pairs of quaternions (a, b) and (x, y). It is straightforward to check that the distributive properties hold and that the necessary properties involving real scalars hold too so this multiplication makes $\mathbb{O}$ into an algebra.

$\mathbb{H} \times \mathbb{H}$ is an eight dimensional real vector space, and the eight pairs

$$O_1 = (1, 0),\, O_2 = (i, 0),\; O_3 = (j, 0),\; O_4 = (k, 0),$$
$$O_5 = (0, 1),\; O_6 = (0, i),\; O_7 = (0, j),\; O_8 = (0, k)$$

span that space. O_1 is the multiplicative identity. For each pair m and n of distinct integers with m and n larger than 1 and not exceeding 8 we find $O_nO_m = -O_mO_n \neq 0$ and also $O_n^2 = -1$.

The octonion multiplication is *not* associative. For instance

$$(O_3O_4)\,O_5 = [\,(j,0)\,(k,0)\,]\ (0,1) = (i,0)\,(0,1) = (0,i) \quad \text{while}$$

$$O_3\,(O_4O_5) = (j,0)\,[\,(k,0)\,(0,1)\,] = (j,0)\,(0,k) = (0,kj) = (0,-i)$$

though it is a fact that associativity *does* hold when restricted to a subspace generated by any *two* octonions.

Failure of associativity means there is no way to represent the octonion structure as a collection of matrices with matrix multiplication.

For octonion q given by

$$q = q_1\,O_1 + q_2\,O_2 + q_3\,O_3 + q_4\,O_4 + q_5\,O_5 + q_6\,O_6 + q_7\,O_7 + q_8\,O_8$$

define the **octonion conjugate** by

$$\tilde{q} = q_1\,O_1 - q_2\,O_2 - q_3\,O_3 - q_4\,O_4 - q_5\,O_5 - q_6\,O_6 - q_7\,O_7 - q_8\,O_8.$$

So if $q = (a,b)$ then $\tilde{q} = (a^*, -b)$ and $q\,\tilde{q} = \tilde{q}\,q = q_1^2 + q_2^2 + \cdots + q_8^2$ corresponds to the square of the usual Euclidean norm on $\mathbb{R}^8$, which we take to be the **norm** on the octonions.

After considerable fussing with terms, one verifies that for two octonions p and q the equality

$$\|\,p\,q\,\| = \|\,p\,\|\,\|\,q\,\|$$

holds: that is, this norm is an algebra norm on the octonions.

We can define the octonion inverse for octonion q to be $\tilde{q}/\,(q\,\tilde{q}) = \tilde{q}/\|\,q\,\|^2$ just as with the quaternions and complex numbers and we see that the octonions form a (non-associative) normed division algebra.

The Cayley-Dickson process can be continued to dimension sixteen and beyond. But the product at the next step (sometimes called the **sedenions**) has zero divisors (nonzero elements h and g with $hg = 0$) and so cannot have an algebra norm and cannot be a division algebra.

20. The Hopf Fibration

This topic has geometrical interest with application to particle physics, but it is not a core idea for us. Still it is a nice idea and illustrates how quaternions help us think about calculations.

We will show that in a certain explicit sense the three dimensional sphere $\mathbb{S}^3$ can be thought of as $\mathbb{S}^2$ copies of the circle $\mathbb{S}^1$ (one circle for each point of the ordinary sphere) glued together.

The structure we consider here is called a **Hopf fibration**, and it is actually realized by a function H called a **Hopf map** from $\mathbb{S}^3$ to $\mathbb{S}^2$ given by the formula

$$H(w,x,y,z) = \left(w^2 + x^2 - y^2 - z^2,\ 2(wz + xy),\ 2(xz - wy)\right).$$

The norm of the output triple is $w^2+x^2+y^2+z^2$ so if (w,x,y,z) is on the unit sphere in $\mathbb{R}^4$ the output is on the unit sphere in $\mathbb{R}^3$, a fact which can be calculated directly or by considering the following.

If we identify a point (w,x,y,z) of $\mathbb{R}^4$ with quaternion $q = w + x\,i + y\,j + z\,k$ and ordered triples in $\mathbb{R}^3$ with the imaginary quaternions we find that

$$q\,i\,q^* = \left(w^2+x^2-y^2-z^2\right)\,i + (2xy+2wz)\,j + (2xz-2wy)\,k$$

so this Hopf map is actually the result of rotating the imaginary i to a new location on the imaginary unit sphere. You can send i anywhere on that sphere by a suitable rotation, so the Hopf map is onto the imaginary unit sphere.

The Hopf map is continuous: if p is close to q then $H(p)$ is close to $H(q)$.

The **fiber over the member g** of the unit imaginary sphere $\mathbb{S}^2$ is the collection of all quaternions on the unit sphere $\mathbb{S}^3$ which move i to g. This collection of unit quaternions will be denoted $\phi_{\mathbf{g}}$ and is nonempty.

If $g = p\,i\,p^* = q\,i\,q^*$ then $p^*\,q\,i = i\,p^*q$; that is, p^*q commutes with i. That will happen if and only if p^*q is of the form $a + b\,i$ where $a^2+b^2=1$ (because both q and p are unit quaternions.)

So if we find any p for which $H(p) = g$ then the fiber ϕ_g over g will be the circle of unit radius inside $\mathbb{S}^3$ given by

$$\phi_g = \{\, p\,(\cos(\theta) + i\,\sin(\theta)) \mid 0 \le \theta < 2\,\pi \,\} = \{\, p\,e^{i\,\theta} \mid 0 \le \theta < 2\,\pi \,\}.$$

These fibers do not overlap: fibers are either equal or disjoint. They form what is called a **partition** of $\mathbb{S}^3$.

We are going to select one member of each fiber.

Define $s(-i) = j$. For other members $g = g_1\,i + g_2\,j + g_3\,k$ of $\mathbb{S}^2$ define $s(g)$ by

$$s(g) = \frac{1}{\sqrt{2(1+g_1)}}\,(\,i\,(1+g_1) + j\,g_2 + k\,g_3\,)\,.$$

We have $s(-i)\,i\,\left(s(-i)\right)^* = j\,i\,(-j) = j\,(-k) = -i$ so j is in the fiber for $-i$.

For other g targets,

$$\begin{aligned} s(g)\,i\,\left(s(g)\right)^* &= \frac{-1}{2(1+g_1)}\,(\,i\,(1+g_1) + j\,g_2 + k\,g_3\,)\;i\;(\,i\,(1+g_1) + j\,g_2 + k\,g_3\,) \\ &= \frac{-1}{2(1+g_1)}\,(\,-(1+g_1) - k\,g_2 + j\,g_3\,)\;(\,i\,(1+g_1) + j\,g_2 + k\,g_3\,) \\ &= i\,g_1 + j\,g_2 + k\,g_3. \quad \left(\text{Multiply and use } 1 - g_1^2 = g_2^2 + g_3^2.\right) \end{aligned}$$

We have now an explicit formula for the fibers:

$$\phi_g = \{\, s(g)\,e^{i\,\theta} \mid 0 \le \theta < 2\,\pi \,\}.$$

Recalling the goal of this section's introductory remarks, we have shown that $\mathbb{S}^3$ can be broken into disjoint copies of $\mathbb{S}^1$ (the fibers) each of which corresponds to one point of $\mathbb{S}^2$.

The function s we used to select a point in each fiber is continuous everywhere except $-i$, the point antipodal to the quaternion i in $\mathbb{S}^2$ which we used to define our fibers.

By choosing any member a of $\mathbb{S}^2$ we could create a new Hopf map and a new collection of fibers consisting of disjoint unit circles in $\mathbb{S}^3$, each circle composed of points p so that $p\,a\,p^* = g$ for each g in $\mathbb{S}^2$. But the selection function analogous to our function s would also fail to be continuous, this time at $-a$.

It is an interesting and important feature of $\mathbb{S}^3$ that no continuous selection of a member of any partition of $\mathbb{S}^3$ into disjoint circles can be found, a fact which leads to the roots of a subject called algebraic topology.

21. (*) Matrices and Quaternions: a Comparison

We will estimate the complexity of doing common rotation calculations in $\mathbb{R}^3$, comparing the efficiency of rotation matrices to quaternions as the "operators" on vectors. We will count the number of real number multiplications required to (i) Rotate a vector. (ii) Calculate an axis and angle of rotation from the operator. (iii) Find the rotation operator given its axis of rotation and angle of rotation. (iv) Combine two operators into a third operator.

We will not count additions or other operations, or data storage requirements, or case-testing operations, or trig function look-ups, or the proliferation of side issues and special cases (all of which hurt the argument for using matrices) so this really is only a rough look at the issue.

(i) Using 3×3 matrix $\mathbf{M}$ on vector v to calculate the rotated vector $\mathbf{M}\,v$ requires 9 multiplications.

Using unit quaternion q to calculate $q\,v\,q^*$ on imaginary v requires two steps. The first, calculate $q\,v$, requires 12 multiplications. Since we know the result of $(q\,v)\,q^*$ will end as imaginary, we don't need to calculate the real term in the second step, so 12 more multiplications are needed.

In view of the formula at the end of Section 17, converting a quaternion to the corresponding matrix operator takes 9 multiplications, so by going *through* the matrix operator we can reduce the number of multiplications to 18.

The score: matrices 9, quaternions 18. Matrices are the big winner here.

(ii) To calculate the axis and angle of rotation from the quaternion operator requires one arccos lookup and no multiplications. If you insist on normalizing the axis vector it requires 6 multiplications and a square root operation.

On the other hand, matrices require more effort. Using the formula at the end of Section 11, we need 3 multiplications and a square root operation and then a trig function lookup to find the angle. To normalize the axis vector requires 3 more multiplications.

So 0 or 6 multiplications for quaternions versus 3 or 6 multiplications for matrices.

(iii) To find the quaternion rotation operator given a normalized axis of rotation and angle of rotation requires two trig function lookups and 3 multiplications. If you need to normalize the axis vector that jumps to 9 multiplications.

To produce the matrix operator, the easiest thing to do is use the formula at the end of Section 17, which requires going *through* the quaternion representation. So there are two trig function lookups and either 3 or 9 multiplications to get there, plus an extra 9 multiplications to create the 9 matrix entries. So it will take 12 or 18 multiplications.

18 versus 9 or 12 versus 3, depending on if you need to normalize the axis vector: quaternions win again.

(iv) To combine two quaternion operators into a third operator requires 16 multiplications. To multiply 2 matrix operators requires 27. Quaternions are better at this.

	Quaternions	**Matrices**
Rotate Vector:	18	9
Find Angle/Axis from Operator:	6 (or 0)	6 (or 3)
Find Operator from Angle/Axis:	9 (or 3)	18 (or 12)
Combine Operators:	16	27

22. (*) Quaternion Interpolation

We saw in Section 21 that quaternions seem to be clearly better at performing two out of four of the tasks we examined, and tied or better at one task, though the analysis was pretty superficial. Matrices came out ahead on one task.

But now we come to a different issue, **interpolating between two operators**, as one would want to do for instance when changing the orientation of a spacecraft or an airplane, or rotating an animated figure in a video game.

The problem with matrix rotation operators is that doing any of the obvious manipulations to transform an initial rotation matrix to a final rotation matrix passes out of the orthogonal matrices, and no inexpensive "normalization" process returns the columns to an orthonormal set of vectors. Rotating one "embedded" axis at a time around predetermined standard axes to its final position (using Euler angles[4], for example) results in jerky and unnatural motion and is subject to a phenomenon called "gimbal lock" where the procedure can break down entirely.

Quaternion interpolation, on the other hand, is relatively easy to describe and understand, and is not subject to these issues, and is numerically stable in the sense that roundoff and other errors remain under control during the interpolation.

And this is why quaternions have become popular in recent years with control theorists (robotics), aeronautical engineers, video software designers and other folks who actually want to *perform calculations* involving objects with changing orientation, as distinct from *proving theorems* about them.

Suppose p_s, p_e, q_s and q_e are unit quaternions, $p_s \neq -p_e$ and $q_s \neq -q_e$. Suppose also that a and b are any continuous increasing real functions defined on the unit interval with $a(0) = b(0) = 0$ and $a(1) = b(1) = 1$.

[4]See **Jack Kuipers**: Quaternions and Rotation Sequences, Princeton University Press, 1999, or **Andrew Hanson**: Visualizing Quaternions, Elsevier, 2006, for more on the various possibilities here.

Then the two quaternion valued functions

$$P(t) = \frac{a(t)\, p_e + (1 - a(t))\, p_s}{\|\, a(t)\, p_e + (1 - a(t))\, p_s \,\|} \quad \text{and} \quad Q(t) = \frac{b(t)\, q_e + (1 - b(t))\, q_s}{\|\, b(t)\, q_e + (1 - b(t))\, q_s \,\|}$$

interpolate continuously between p_s and p_e, and q_s and q_e, respectively, in the unit quaternions. (The most obvious choices for a and b are $a(t) = b(t) = t$.)

Specifically,

$$\|P(t)\| = \|Q(t)\| = 1 \text{ for all } t \text{ in } [0, 1]$$

and

$$P(0) = p_s, \quad P(1) = p_e, \quad Q(0) = q_s, \quad \text{and} \quad Q(1) = q_e.$$

P and Q are confined to the subspaces of $\mathbb{H}$ generated by p_s and p_e and by q_s and q_e, respectively.

If v is an imaginary quaternion and w is any quaternion then

$$P(t)\, v\, P(t)^* \quad \text{and} \quad P(t)\, w\, Q(t)$$

are continuous paths in the quaternions, starting at $p_s\, v\, p_s^*$ and $p_s\, w\, q_s$ and ending at $p_e\, v\, p_e^*$ and $p_e\, w\, q_e$, respectively. The norms of these quaternions remain constant along the path. And, of course, these quaternions correspond to vectors rotating from one position to the other in $\mathbb{R}^3$ and $\mathbb{R}^4$, respectively.

A **discrete approximation** to this continuous model of rotation would be obtained by selecting times t_i in [0, 1], $i = 0, \ldots, n$ with

$$0 = t_0 < t_1 < t_2 < \cdots < t_n = 1$$

for which consecutive $P(t_i)$ and $Q(t_i)$ are sufficiently close together for the purposes at hand. Using the method of interpolation outlined above, guarantee of "sufficiently close" may be an issue.

Even so, this simple formula has no counterpart that can be discussed in 4 or 5 paragraphs—or 4 or 5 pages—using matrices without reprising the ideas we have introduced involving quaternions. The option of creating a linear interpolation between starting matrix and ending matrix, and performing Gram-Schmidt orthonormalization at each step from start to finish is computationally unappealing.

The "video animation community" refers to these quaternion interpolation techniques by the quasi-acronym **slerp**, short for **spherical linear interpolation,** in reference to the fact that the quaternions that implement these rotations lie on the unit sphere $\mathbb{S}^3$ in $\mathbb{H}$.

There is another way of interpolating between unit quaternions that has some serious advantages.

If p_s and p_e are unit quaternions, so is $h = p_s^*\, p_e$.

In polar form, $h = \cos(\theta) + \sin(\theta)\, \mathbf{U}$ for argument θ and unit imaginary $\mathbf{U}$, and $\mathbf{U}$ can be chosen so that $0 \leq \theta \leq \pi$.

So $\log(h) = \ln(\|h\|) + \theta\, \mathbf{U} = \theta\, \mathbf{U}$.

The function

$$R(t) = p_s\, (\, p_s^*\, p_e\,)^t = p_s \left(h^t\right) \quad \text{for } t \in [0, 1]$$

satisfies $R(0) = p_s$ and $R(1) = p_e$.

Also, since $h^t = e^{t\,\log(h)} = e^{t\,\theta\,\mathbf{U}} = \cos(t\,\theta) + \sin(t\,\theta)\mathbf{U}$ we have

$$R(t) = p_s\left(h^t\right) = p_s\left(\cos(t\,\theta) + \sin(t\,\theta)\,\mathbf{U}\right).$$

It is apparent that $R(t)$ interpolates from p_s to p_e through *unit* quaternions.

The restriction that $p_s \neq -p_e$ is lifted: $\theta = \pi$ and $\mathbf{U}$ can be any choice of unit imaginary in this case.

This interpolation proceeds at constant speed $|\theta|$ so for each i

$$\|R\left(t_{i+1}\right) - R\left(t_i\right)\| \ \leq\ |\theta|\,|\,t_{i+1} - t_i\,|\,.$$

23. (*) The "Belt Trick"

The fact that each rotation R in $\mathbb{R}^3$ can be implemented by two different unit quaternions q and $-q$ through

$$R(v) = q\,v\,q^* \quad \text{and} \quad R(v) = (-q)\ v\ (-q^*)$$

has interesting and, at first glance, rather puzzling consequences both sublime (particle physics) and mundane (macroscopic objects.)

Let's suppose that we are going to rigidly rotate the standard coordinate basis in $\mathbb{R}^3$, passing through intermediate steps in a continuous way, and after a period of time, say 1 minute, we end up back where we started: at the standard coordinate basis.

We envision this process to be determined by a continuous unit quaternion valued function T_t for t in $[0,1]$ where

$$T_t = \cos\left(\theta_t\right) + \sin\left(\theta_t\right)\ \mathbf{U}_t$$

for continuous real function θ_t and continuous unit imaginary quaternion valued function $\mathbf{U}_t$ defined, themselves, on the unit interval.

So the continuous rotation we intend is implemented by

$$R_t(v) = T_t\,v\,T_t^* \quad \text{for } v \text{ in } \mathbb{R}^3 \text{ and } t \text{ in } [0,1].$$

Given our conditions both R_0 and R_1 must be the identity map: no visible rotation from the standard basis at all. We may choose θ_0 to be 0 and T_0 to be $1 = \cos(0) + \sin(0)\ \mathbf{U}_0$. But then there are two possibilities for T_1, namely

$$T_1 = 1 = 1 + 0\,\mathbf{U}_1 \quad \text{or} \quad T_1 = -1 = -1 + 0\,\mathbf{U}_1.$$

The first case corresponds to θ_t ending at an even multiple of π while the second case has θ_t ending at an odd multiple of π.

Remember that a unit quaternion at angle θ causes a rotation by angle $2\,\theta$ around its imaginary part, so in both cases we end up at a rotation by an even multiple of π, both apparently equivalent (at least to the eye) to no rotation at all.

In the first case, using the interpolation techniques of the last section for each fixed t if you wish, *the path through the unit quaternions represented by T can be continuously deformed to the constant function* 1 *while leaving both of the endpoints, T_0 and T_1, invariant. In the second case this cannot be done.*

The remark concerning the second case is obvious: you cannot continuously change -1 to 1 (or 1 to -1) without leaving the quaternions that correspond to the identity, thereby violating the condition that $R(0) = R(1)$ and that both remain the identity during the transformation.

Here is a simple physical scenario that demonstrates the difference between the two cases: the **belt trick**. There are several variations, based on the same principle.

Take a belt and nail its tip to the floor (or step on it.) Grasp the belt and pull it straight so you are looking down along its length to the floor, your eye gazing upon the flat surface of the belt. Imagine many parallel copies of the three coordinate axes running up the spine of the belt with each y axis pointing generally up towards you as you look down the length of the belt from above, each z axis pointing out and away from the surface of the belt and each x axis pointing to the left.

Now give the belt a $720°$ twist along its length.

The axes affixed to the belt at various places along its length are rotated around the y axis from their original position by various amounts, but the axes at the floor and the axes gripped between your fingers at the top are in their original position.

The amount of rotation along the belt is a function of position along the belt and represents a path through the unit quaternions that starts and ends at the unit quaternion 1.

Moving around in various ways but *not rotating the axes in your fingers or those at the floor* you can straighten the belt.

Try it yourself, and try to visualize the changing path through the unit quaternions (as a function of position along the length of the belt) as it transforms from the unit circle in the $1, j$ plane to the constant 1 function.

If the belt is twisted only $360°$ you cannot untwist the belt, no matter what you do, without rotating the axes at the floor or in your fingers or tearing the belt.

An argument of this kind is called *topological* and using it we drew powerful physical conclusions without doing complicated calculations.

Index