Complex numbers and quaternions*

Timothy Poston

CIeMed, ISS, NUS

February 17, 1998

Special Effects in the next "Star Trek"

In motion pictures (such as Lawnmower Man, Looker, Return of the Jedi, and the Genesis effect in Star Trek 2), computer graphics effects have been used quite successfully to depict, um, computer graphics. Now, the artifacts of computer graphics will be used to best advantage in

Star Trek NaN : The Wrath of Phong

The landing party alpha-blends to opacity on the planet's surface.

SPOCK : The tricorder indicates a complete lack of mirror reflections and cast shadows. Also, if we travel too far from the origin, we will suffer from round-off error. Captain, this planet is highly dangerous; it was wise to bring expendable red-shirted security men.

An expendable red-shirted security man turns to the side while looking upward. His neck joint suffers gimbal lock, and he falls in a heap. The group rushes to him.

MCCOY : Jim, he's dead.

KIRK : *throws arms wide, ripping shirt* No quaternions? What kind of planet is this?

1 2×2 Matrices

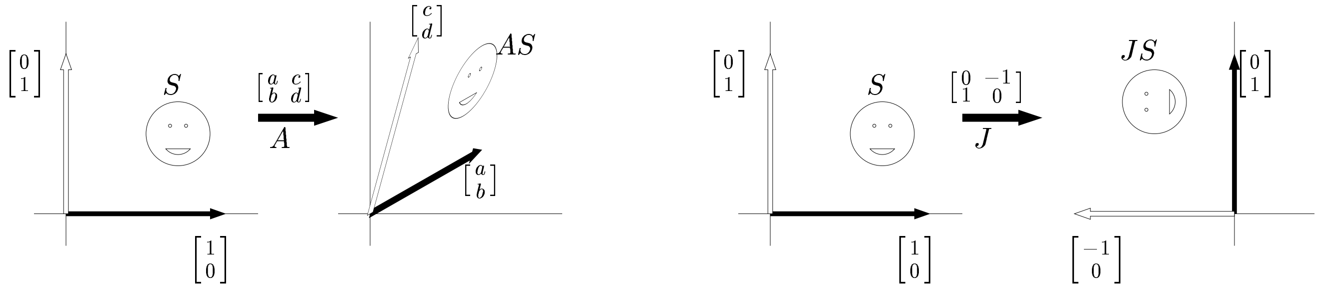

Figure 1. The matrix for A is a list of column vectors, telling you how to draw a picture of where each basis vector (in the basis we're currently using) goes to under the action of A.
In particular, rotation J through 90° takes $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ to $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ to $\left[\begin{smallmatrix}-1\\0\end{smallmatrix}\right]$, so its matrix is $\left[\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right]$. The identity I takes $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ to $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ to $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ (of course!), so its matrix is $\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]$.

*With thanks to Lim, Ik Soo for very careful reading, many corrections of ± signs, and much painstaking work with the figures.

To make geometric sense of quaternions, start with complex numbers. To make geometric sense of those, see them dynamically, in terms of real linear operations, rather than as a mystery introduced by mad algebraists to give equations like $x^2+1=0$ solutions they seem not to have. 'Imaginary numbers' were called 'impossible numbers' for the first century or so, until the introduction of the geometric 'complex plane' viewpoint calmed everybody down. But even that is *static* geometry. I prefer to think of them dynamically.

To do that, start by thinking of linear transformations of the ordinary real (x, y)-plane: every vector $\mathbf{p}$ goes to a vector $A\mathbf{p}$, and we have $A(\mathbf{p}+\mathbf{q}) = A\mathbf{p} + A\mathbf{q}$ and $A(s\mathbf{p}) = sA\mathbf{p}$. Then if we know that $A\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] = \left[\begin{smallmatrix}a\\b\end{smallmatrix}\right]$ and $A\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right] = \left[\begin{smallmatrix}c\\d\end{smallmatrix}\right]$, by linearity we know that

$$\begin{aligned} A\begin{bmatrix}x\\y\end{bmatrix} &= A\left(x\begin{bmatrix}1\\0\end{bmatrix}+y\begin{bmatrix}0\\1\end{bmatrix}\right) = A\left(x\begin{bmatrix}1\\0\end{bmatrix}\right)+A\left(y\begin{bmatrix}0\\1\end{bmatrix}\right) = xA\begin{bmatrix}1\\0\end{bmatrix}+yA\begin{bmatrix}0\\1\end{bmatrix} \\ &= x\begin{bmatrix}a\\b\end{bmatrix}+y\begin{bmatrix}c\\d\end{bmatrix} = \begin{bmatrix}ax\\bx\end{bmatrix}+\begin{bmatrix}cy\\dy\end{bmatrix} = \begin{bmatrix}ax+cy\\bx+dy\end{bmatrix} \end{aligned} \tag{1.1}$$

which is where the usual formula for matrix multiplication

$$\begin{bmatrix}a & c\\b & d\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}ax+cy\\bx+dy\end{bmatrix} \tag{1.2}$$

really comes from: think of $\left[\begin{smallmatrix}a & c\\b & d\end{smallmatrix}\right]$ as a list of the column vectors $\left[\begin{smallmatrix}a\\b\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}c\\d\end{smallmatrix}\right]$ that $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ go to, and formula (1.2) follows immediately. So does the 'matrix multiply' formula for the combined operation C defined by "do A then B", which we write as $C = BA$ because "B of (A of $\mathbf{p}$)" is written as $B(A\mathbf{p})$ or $BA\mathbf{p}$. Just apply B, using (1.2), to the column vectors of A, which are the vectors that A takes $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ to. This gives the vectors that "A then B" takes $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ to; that is, the column vectors of C.

We can also look at what A does to *shapes*, like the smiley S in Fig. 1. By mapping each point $\mathbf{p}$ in S to $A\mathbf{p}$, we build up a new set of points; call it AS. In general the shape of AS changes in a complicated way from the shape of S. Some lengths may become longer, some shorter, some angle widen while others narrow, and so on. But *areas* behave very simply. They are all multiplied—independent of shape—by the same number, which we call $\mathsf{det}\, A$. By definition, then, $\text{area}(AS) = \mathsf{det}\, A\, \text{area}(S)$ for any S. (We do need to think of area as having a sign, or handedness, just as length needs a sign for '10 miles travel' to add correctly to '5 miles travel' when the 10 is towards Moscow from Kiev and the 5 is toward Kiev. With 15 miles of wear on shoe leather we are only 5 miles nearer Moscow. Flipping direction or handedness changes sign, and det reflects this.)

Seeing $\mathsf{det}\, A$ as dynamic, as the area multiplication associated with A, makes it clear that when C means "do A then B" its effect on areas must be the same as first multiplying by $\mathsf{det}\, A$, then multiplying by $\mathsf{det}\, B$ ... or multiplying once by their product. In other words,

$$\text{if} \quad C = BA\,, \quad \text{then} \quad \mathsf{det}\, C = \mathsf{det}\, B\, \mathsf{det}\, A\,, \tag{1.3}$$

a fact which is totally unintelligible if you start with a *formula* for det instead of a *concept* of it. (The formula is easily derived from the concept, even in the $n\times n$ case, where 'area' is replaced by 'n-volume'.)

2 Complex numbers

Complex numbers correspond to a rotation-flavoured slice of the collection of all real linear operators on the plane. To write a matrix as

$$A = \begin{bmatrix} a & c \\ b & d \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + d\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.1}$$

is simply to describe it in terms of the basis vectors $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]$. But this choice of basis is an awkward one; for instance, the identity $I = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ is a combination of two of them. The identity operator is so special (having the same matrix in all basis systems for the plane) that we do well to use it as a basis vector, as in

$$\begin{aligned} A = \begin{bmatrix} a & c \\ b & d \end{bmatrix} &= \frac{a+d}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{b-c}{2}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \frac{a-d}{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \frac{b+c}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ &= \alpha\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \beta\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \mu\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \nu\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \text{ say.} \qquad (2.2) \\ &= \begin{bmatrix} \alpha+\mu & \nu-\beta \\ \beta+\nu & \alpha-\mu \end{bmatrix} \end{aligned}$$

It is easy to pick out the 'rotation flavoured' operators like $\left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$ from reflections like $\left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$; use the determinant, which is positive if A does not turn the plane over. We have

$$\mathsf{det}\, A = \alpha^2 + \beta^2 - \mu^2 - \nu^2 \,, \tag{2.3}$$

easily checked by multiplying out $\left(\frac{a+d}{2}\right)^2 + \left(\frac{b-c}{2}\right)^2 - \left(\frac{a-d}{2}\right)^2 - \left(\frac{b+c}{2}\right)^2$. So 'all the A of the form $\alpha\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right] + \beta\left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$', or, more tidily, 'of the form $\alpha I + \beta J$' gives a plane where $\mathsf{det} = \alpha^2 + \beta^2$, positive for all non-zero A. (Any 3D subspace would include an A with $\mathsf{det}\, A < 0$, so this is as big a 'positive' subspace as there is.) Since J is a 90° turn, $J^2 = -I$ as a 180° turn carries every $\mathbf{p}$ to $-\mathbf{p}$, and so

$$\begin{aligned} \big(\alpha I + \beta J\big)\big(\widehat{\alpha} I + \widehat{\beta} J\big) &= \alpha\widehat{\alpha} I^2 + \beta\widehat{\alpha} JI + \alpha\widehat{\beta} IJ + \beta\widehat{\beta} J^2 && (2.4) \\ &= \big(\widehat{\alpha}\alpha - \beta\widehat{\beta}\big)\, I + \big(\alpha\widehat{\beta} + \beta\widehat{\alpha}\big)\, J\,. && (2.5) \end{aligned}$$

Denote I by the name 1 (it is the same as *multiplication* by 1 —think of numbers by what they do, rather than what they 'are'), and J by i; we have $i^2 = -1$ without any mysticism. The symbols i and -1 represent actions ('quarter turn' and 'about face'), not 'counting numbers': squaring an action is taking it twice: doing i twice is the same as doing -1 once. We *call* the system of things like $\alpha I + \beta J$, with its multiplication and addition, the 'complex numbers', but they generalize number operations like trebling and halving, not the primary school one of 'how many toes?'.

3 2D Rotations

A rotation by θ takes $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ to $\left[\begin{smallmatrix} \cos\theta \\ \sin\theta \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ to $\left[\begin{smallmatrix} -\sin\theta \\ \cos\theta \end{smallmatrix}\right]$, so its matrix is

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \cos\theta\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \sin\theta\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \cos\theta + i\sin\theta\,, \text{ as a complex number.} \tag{3.1}$$

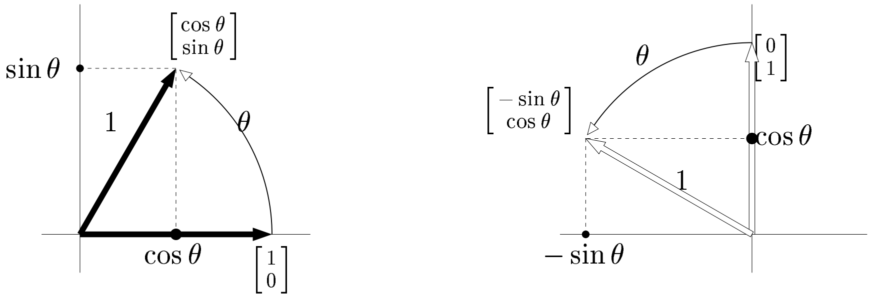

Figure 2. Rotating the standard basis vectors by θ.

The collection of all $A = \alpha + i\beta$ that are rotations is the circle $\mathsf{det}\, A = \alpha^2 + \beta^2 = 1$ in the complex plane. Any complex number $x + iy$ is just some r times some $\cos\theta + i\sin\theta$, where r is the distance $\sqrt{x^2 + y^2}$ and θ is the angle $\arctan2(y, x)$ that $x + iy$ makes with the line through 1 (the 'positive real axis'). Multiplying by a real number commutes with any matrix (by linearity), so

$$
\begin{aligned}
(x+iy)(\widehat{x}+i\widehat{y}) &= \left(r\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\right)\left(\widehat{r}\begin{bmatrix}\cos\widehat{\theta} & -\sin\widehat{\theta}\\ \sin\widehat{\theta} & \cos\widehat{\theta}\end{bmatrix}\right)\\
&= r\widehat{r}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}\cos\widehat{\theta} & -\sin\widehat{\theta}\\ \sin\widehat{\theta} & \cos\widehat{\theta}\end{bmatrix} \qquad (3.2)\\
&= r\widehat{r}\left(\text{rotate by } \theta \text{ after rotating by } \widehat{\theta}\right)\\
&= r\widehat{r}\left(\text{rotate by } \theta+\widehat{\theta}\right)\\
&= r\widehat{r}\left(\cos(\theta+\widehat{\theta}) + i\sin\left(\theta+\widehat{\theta}\right)\right) \qquad (3.3)
\end{aligned}
$$

(Writing out (3.2) as a matrix multiplication gives $\cos(\theta+\widehat{\theta})$ and $\sin(\theta+\widehat{\theta})$ in terms of $\cos\theta$, *etc.*: I could never remember those 'trig formulæ' until I realized that.) Doing one rotation after another—which is what matrix multiplication represents—corresponds to adding the angles turned through. So multiplying $x+iy$ and $\widehat{x}+i\widehat{y}$ multiplies their lengths while adding their angles, though that would be an inefficient way to compute $(x+iy)(\widehat{x}+i\widehat{y})$. It is better to use complex algebra as a sneaky way to do trigonometry and 2D rotations.

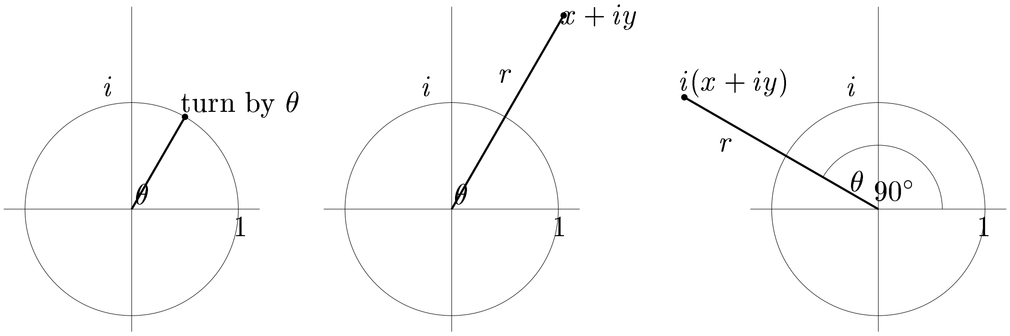

Figure 3. If $(x+iy)$ is r[turn by θ], then $i(x+iy)$ is [turn by $90°$] r[turn by θ]; multiplication by $i = \left[\begin{smallmatrix}0 & -1\\ 1 & 0\end{smallmatrix}\right]$ turns $x+iy = \left[\begin{smallmatrix}x & -y\\ y & x\end{smallmatrix}\right]$ through a right angle, just like the vectors $\left[\begin{smallmatrix}u\\ v\end{smallmatrix}\right]$ that $\left[\begin{smallmatrix}0 & -1\\ 1 & 0\end{smallmatrix}\right]$ was defined to act on.

4 Exponentials

The next thing to look at is complex exponentials. "The number e multiplied by itself $\left[\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right]$ times" makes very little sense. The trick is (a) to think of numbers as (multiplicative) *actions*, (b) to compound those actions, taking money and interest as an example. In its simplest version, "annual interest rate of 10%" means "add $P/10$ to the current principal P, once a year," which means "replace P by $P + P/10$" or "multiply P by $1+\frac{1}{10}$".

Compound interest over N years means N multiplications by $1+\frac{1}{10}$ or—cumulatively—multiplication by $\left(1+\frac{1}{10}\right)^N$, which goes 1, 1.1, 1.21, 1.331, 1.4641, $\ldots$. Annual interest of 100% (the simple 'loan shark' rate), is annual doubling, 1, 2, 4, 8,$\ldots$.

What if you collect part of the interest early, and reinvest? In half a year, at the same 100% annual rate, you add $P/2$; that is, half a year multiplies by $(1 + 1/2)$. Doing it twice multiplies P by $(1 + 1/2)^2 = 2.25$ over the year, not just by 2. Compounding every three months multiplies by $1 + 1/4$ each time, by $\left(\frac{5}{4}\right)^4 = 2.44140625$ over the year, and so on. Over time t, divided into N steps of t/N, we multiply by $(1 + t/N)^N$. As N increases, so does this multiplier, but although it always get larger it has an upper limit. Let us denote value that $(1 + t/N)^N$ creeps upward to by $E(t)$:

$$E(t) = \lim_{N\to\infty} \left(1 + \frac{t}{N}\right)^N , \tag{4.1}$$

the multiplier for arbitrarily-often compounding at rate 1 (alias cent per cent) for time t. From the definition (though less obviously from (4.1)) it seems reasonable that

$$\begin{aligned} E(t+\tau) &= \text{multiplier for time } t+\tau & (4.2)\\ &= (\text{multiplier for time } t) \times (\text{multiplier for time } \tau) , & (4.3) \end{aligned}$$

because we sit separately and in order through the times t and τ. Thus

$$E(t+\tau) = E(t)E(\tau) . \tag{4.4}$$

In particular

$$\begin{aligned} E(nt) &= E(t+t+\cdots+t) &= E(t)E(t)\ldots E(t) &= (E(t))^n , & (4.5)\\ \left(E(\tfrac{1}{2}t)\right)^2 &= \left(E(\tfrac{1}{2}t)\right)\left(E(\tfrac{1}{2}t)\right) &= E\left(\tfrac{1}{2}t+\tfrac{1}{2}t\right) &= E(t) , & (4.6)\\ \text{so } E(\tfrac{1}{2}t) &= \sqrt{E(t)} &= (E(t))^{\frac{1}{2}} , & & (4.7)\\ \text{and in general } E(st) &= (E(t))^s . & & & (4.8) \end{aligned}$$

In particular, then,

$$E(t) = E(t1) = (E(1))^t . \tag{4.9}$$

We give $E(1)$, whose value turns out to be approximately $2.7182818\ldots$, the name e, so the multiplier for 'continuously compounding over time t at cent per cent' gets the formula e^t. The fundamental point, though, is the continuous compounding: not 'multiplying e by itself

t times', which is already a peculiar idea if t is a number like $\sqrt{2}$, but the effect of growing a steady multiplicative rate for time t. In the general case of an annual rate R (alias $100R\%$), we get a multiplier $(1+R\tau)$ in a time step of τ, so in N steps of time t/N we get a multiplier $(1+R(t/N))^N$, and going to the limit a multiplier

$$\lim_{N\to\infty}\left(1+R\frac{t}{N}\right)^N = \lim_{N\to\infty}\left(1+\frac{Rt}{N}\right)^N = e^{Rt} \text{ by definition.} \tag{4.10}$$

We have constructed a solution e^{Rt} for $dP/dt = RP$, the equation for growth at proportional rate R, by looking at the details of the proportional growth process.

So what happens when $R = i$?

From the differential equation point of view, solving $dz/dt = iz$, the result is geometrically obvious.

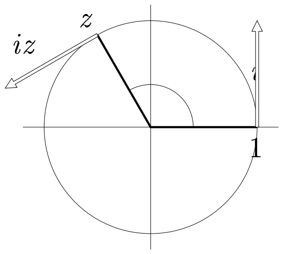

Figure 4. The differential equation $dz/dt = iz$ says that z must be changing with velocity iz, which is at right angles to z.

This fits with $z(t)$ going around in circles, which we can write down and test; the formula for that is $z(t) = r(\cos t + i\sin t)$ if it starts on the real axis, and this works:

$$\frac{d}{dt}\Big(r(\cos t + i\sin t)\Big) = r\Big(-\sin t + i\cos t\Big) = i\Big(r(\cos t + i\sin t)\Big)\ , \tag{4.11}$$

so d/dt of this $z(t)$ is indeed i times this $z(t)$, but let's see it geometrically (Fig. 5).

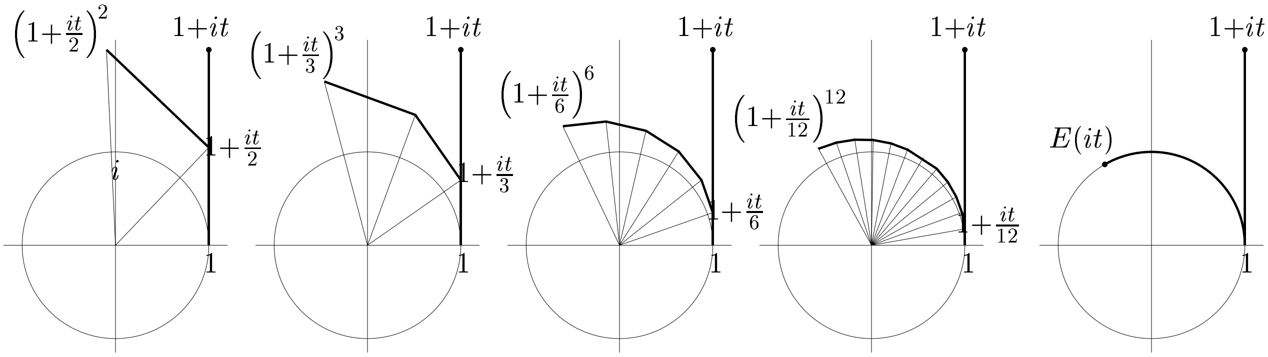

Figure 5. Multiplying by $\left(1+\dfrac{it}{N}\right)$ is a stretch by $\sqrt{1+\dfrac{t^2}{N^2}}$, a turn by the angle that $\left(1+\dfrac{it}{N}\right)$ makes with the real axis. As N grows, the N-fold result wraps closer around the unit circle; in the limit $E(it) = \cos t + i\sin t$. Equal lengths of the line are mapped to equal lengths of the circle, because smaller pieces look more like the multiplication step from 1 to $1+it/N$, which is more and more like a tiny piece of the circle as N grows.

We extend the notation e^{whatever} to this case:

$$e^{it} = \cos t + i\sin t \tag{4.12}$$

is a simple consequence of complex multiplication going round and round, since unit complex numbers are rotations; $1 + it$ is 'nearly a rotation' if we think of the it as an 'infinitesimal rotation'. The equation

$$e^{i\pi} = -1 \tag{4.13}$$

just says that -1 and π are two different labels (one the shortest way, one the prettiest) for 'halfway around the unit circle'.

Now, what is this complex $E(\,)$, alias $\exp(\,)$ or 'e to the ...', really defined on? The values lie in the circle S^1 of unit complex numbers—which is really the *group* of real 2D rotations—but it is really the tangent line at 1 to S^1 that is really being wrapped around S^1. We write $E(it)$, as though we were evaluating $E()$ on the complex number it, but what we put together to find it are repetitions of multiplication by $1 + it/N$, not by it/N. Vectors like it/N are often called 'infinitesimal rotations', but it is only by basing them at 1 that this makes sense; it is $1 + it/N$ that is near to being a small rotation—turn it by just a little—in the ordinary sense. Of course, it is on the line, not the circle, but the smaller it/N is, the smaller *even in proportion to* it/N is the error in using $1 + it/N$ as an approximation to a genuine rotation. Conversely, the best 'difference' in Fig. 6 between the rotation $R_\theta = \cos\theta + i\sin\theta$ and the rotation $R_{\theta+\delta}$ is *not* the difference vector

$$\mathbf{v} = R_{\theta+\delta} - R_\theta = \cos(\theta+\delta) - \cos\theta + i\left(\sin(\theta+\delta) - \sin\theta\right)\ ; \tag{4.14}$$

it is the small rotation $R_\delta = R_{\theta+\delta}R_{-\theta}$, near 1, which we can combine with R_θ to give $R_{\theta+\delta}$. This is the difference that belongs to the multiplication structure of the complex numbers, which comes from their nature as rotations, not the one that belongs to the addition structure.

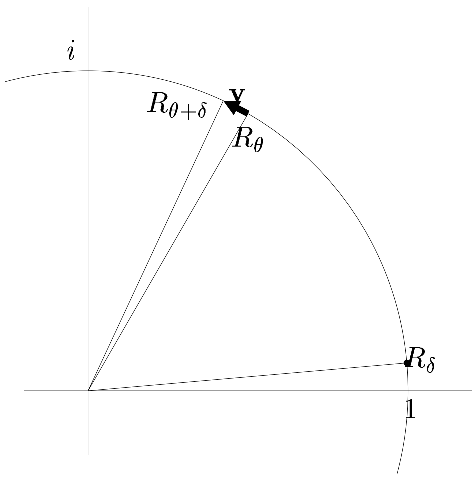

Figure 6. What is the 'small change' between rotation R_θ through θ and rotation $R_{\theta+\delta}$ through $\theta+\delta$?

If we have a time-dependent rotation $R_{\theta(t)}$, its rate of change is similarly an 'infinitesimal rotation' near 1, represented by a vector starting at 1:

$$\frac{d}{dt}e^{i\theta(t)} = \left(i\frac{d\theta}{dt}\right)e^{i\theta(t)}\,. \tag{4.15}$$

This gets all rotational velocities into a common space, tangent to the circle at 1, instead of scattered all over the group. This is handy for jobs like filtering and predicting (if my rotational velocity is ω, then to first order

$$R_{\theta(\text{now}+t)} = \omega t \, R_{\theta(\text{now})} \tag{4.16}$$

and so on.)

Of course, for 2D rotations θ labels orientations perfectly smoothly (as long as we remember that $\theta + 2\pi = \theta$), and could use $d\theta/dt$ as a rate of turn. In 3D, life is far more complicated. A rotation is given by a 3×3 matrix $M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$, and the space of all these, labelled by the m_{ij}, is nine-dimensional. The matrices that represent actual rotations must satisfy the six independent equations[1]

$$\begin{aligned}
m_{11}^2 + m_{21}^2 + m_{31}^2 &= 1 && (4.17)\\
m_{12}^2 + m_{22}^2 + m_{32}^2 &= 1 && (4.18)\\
m_{13}^2 + m_{23}^2 + m_{33}^2 &= 1 && (4.19)\\
m_{11}m_{12} + m_{21}m_{22} + m_{31}m_{32} &= 0 && (4.20)\\
m_{11}m_{13} + m_{21}m_{23} + m_{31}m_{33} &= 0 && (4.21)\\
m_{12}m_{13} + m_{22}m_{23} + m_{32}m_{33} &= 0 && (4.22)
\end{aligned}$$

which restricts them to a merely 3D set, just as a curve in 3-space can be given (Fig. 7) by two equations $x^2 + y^2 + z^2 = 1$ and $x^2 - y^2 + z = 0$, leaving us with a single dimension along the curve. It is a richly twisted 3D set in 9D; like the curve in Fig. 7, it is not restricted to any flat subset of the surrounding space. The set {90° and 180° turns about x, y and z axes, 180° turns about the diagonals of the (x, y), (y, z) and (x, z) planes} gives a basis for the 9D space of all 3×3 matrices. There is no *linear* equation obeyed by all rotations in the way they obey the six quadratic ones (4.17-4.22).

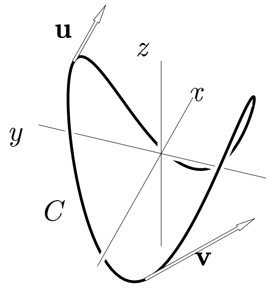

Figure 7. The set C of points with $x^2 + y^2 = 1$ and $x^2 - y^2 + z = 0$ is only 1D, but it sticks out in 3D. Tangent vectors like **u** and **v** to it at different points can be added to yield *any* 3D vector.

Moreover, no system of angles works to label 3D rotations well. It is a topological theorem that *any* labelling scheme goes bad somewhere on the group, like Longitude at the Earth's poles. This has horrible consequences of noisy behaviour from a 3D sensor near the corresponding positions, if you communicate orientations by a scheme like Euler angles, as

[1] And one inequality, $\det M > 0$, to separate them from the reflections. Reflections satisfy (4.17-4.22), but have determinant -1.

experience in the ISS VR lab illustrates. Consequently, since we cannot compare velocity vectors at different points on the 3D rotation group by treating them as rates of change of labels in a nice three-angle labelling scheme, and since they scatter over all nine dimensions' worth of directions—in different 3D sets of directions at each point—it is essential to pull them back to a common space, the tangent vectors (infinitesimal rotations) at the 3D identity operator I; tangent to the group at I, not at the current point in the group. These turn out to be the skew-symmetric matrices. When the rotational velocity is the skew-symmetric matrix A, the predicted rotation in time t is e^{At} (constructed just like the e^{it} above), and so on; but e^{At} is a twisty map, to match the twisty set where its values live. If we use an approximation to it, getting 'not quite a rotation', it is a subtle matter to find the nearest true rotation. All this becomes complicated to think about, and CPU-costly to implement.

We can get back to geometry and calculations that are almost as simple as the 2D case, by repeating the same algebra as above, this time for *complex* 2×2 matrices. What we find is the quaternions.

5 Quaternions

Take the general 2×2 complex matrix $A = \left[\begin{smallmatrix} a+ib & e+if \\ c+id & g+ih \end{smallmatrix}\right]$ and arrange its determinant, just as in (2.3), as a sum and difference of squares. We have $\mathsf{det}\,A$ as

$$\begin{aligned}
\mathsf{det}\begin{bmatrix} a+ib & e+if \\ c+id & g+ih \end{bmatrix} &= (a+ib)(g+ih)-(c+id)(e+if) \\
&= ag-bh-ce+df+i(bg+ah-cf-de) \\
&= \left(\frac{a+g}{2}\right)^2+\left(\frac{c-e}{2}\right)^2+\left(\frac{b-h}{2}\right)^2+\left(\frac{d+f}{2}\right)^2 \\
&\qquad -\left(\frac{a-g}{2}\right)^2-\left(\frac{b+h}{2}\right)^2-\left(\frac{c+e}{2}\right)^2-\left(\frac{d-f}{2}\right)^2+i\big(\text{stuff}\big) \\
&\equiv \alpha^2+\beta^2+\gamma^2+\delta^2-\mu^2-\nu^2-\sigma^2-\rho^2+i\big(\text{stuff}\big)\ , \text{ say,}
\end{aligned} \tag{5.1}$$

where α *etc.* are defined by the respective ( )s. This corresponds to writing A as

$$\alpha\begin{bmatrix}1&0\\0&1\end{bmatrix}+\beta\begin{bmatrix}0&-1\\1&0\end{bmatrix}+\gamma\begin{bmatrix}i&0\\0&-i\end{bmatrix}+\delta\begin{bmatrix}0&i\\i&0\end{bmatrix}+\mu\begin{bmatrix}1&0\\0&-1\end{bmatrix}+\nu\begin{bmatrix}i&0\\0&i\end{bmatrix}+\sigma\begin{bmatrix}0&1\\1&0\end{bmatrix}+\rho\begin{bmatrix}0&-i\\i&0\end{bmatrix}+i\left(\begin{smallmatrix}\text{matrix}\\\text{stuff}\end{smallmatrix}\right)$$

$$=\begin{bmatrix}\alpha&0\\0&\alpha\end{bmatrix}+\begin{bmatrix}0&-\beta\\\beta&0\end{bmatrix}+\begin{bmatrix}i\gamma&0\\0&-i\gamma\end{bmatrix}+\begin{bmatrix}0&i\delta\\i\delta&0\end{bmatrix}+\begin{bmatrix}\mu&0\\0&-\mu\end{bmatrix}+\begin{bmatrix}i\nu&0\\0&i\nu\end{bmatrix}+\begin{bmatrix}0&\sigma\\\sigma&0\end{bmatrix}+\begin{bmatrix}0&-i\rho\\i\rho&0\end{bmatrix}+i\left(\begin{smallmatrix}\text{matrix}\\\text{stuff}\end{smallmatrix}\right)$$

to give $\mathsf{det}\,A$ the simple form (5.1). As before, we pick out those components that make a positive contribution to the determinant; this time, positive *real*. That is, we look at matrices of the form

$$A = \alpha\begin{bmatrix}1&0\\0&1\end{bmatrix}+\beta\begin{bmatrix}0&-1\\1&0\end{bmatrix}+\gamma\begin{bmatrix}i&0\\0&-i\end{bmatrix}+\delta\begin{bmatrix}0&i\\i&0\end{bmatrix} \tag{5.2}$$

$$= \alpha+i\beta+j\gamma+k\delta \tag{5.3}$$

where 1, i, j, k stand for $\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}i&0\\0&-i\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0&i\\i&0\end{smallmatrix}\right]$ respectively. (These also go by the name of 'the **Pauli spin matrices**'.)

These $A = \alpha + i\beta + j\gamma + k\delta$ are the **quaternions**, equipped with the addition and multiplication they have as matrices. This multiplication does *not* commute, obeying $ab = ba$ the way real and complex numbers do. As for any matrices, addition commutes, but simple matrix multiplication gives

$$ij = \begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{bmatrix}i&0\\0&-i\end{bmatrix} = \begin{bmatrix}0&i\\i&0\end{bmatrix} = k \tag{5.4}$$

$$ji = \begin{bmatrix}i&0\\0&-i\end{bmatrix}\begin{bmatrix}0&-1\\1&0\end{bmatrix} = \begin{bmatrix}0&-i\\-i&0\end{bmatrix} = -k \tag{5.5}$$

$$jk = \begin{bmatrix}i&0\\0&-i\end{bmatrix}\begin{bmatrix}0&i\\i&0\end{bmatrix} = \begin{bmatrix}0&-1\\1&0\end{bmatrix} = i \tag{5.6}$$

$$kj = \begin{bmatrix}0&i\\i&0\end{bmatrix}\begin{bmatrix}i&0\\0&-i\end{bmatrix} = \begin{bmatrix}0&1\\-1&0\end{bmatrix} = -i \tag{5.7}$$

$$ki = \begin{bmatrix}0&i\\i&0\end{bmatrix}\begin{bmatrix}0&-1\\1&0\end{bmatrix} = \begin{bmatrix}i&0\\0&-i\end{bmatrix} = j \tag{5.8}$$

$$ik = \begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{bmatrix}0&i\\i&0\end{bmatrix} = \begin{bmatrix}-i&0\\0&i\end{bmatrix} = -j\,. \tag{5.9}$$

Quaternion multiplication does not commute, and this means that many of the features of 'numbers' shared by whole, real and complex numbers break down here. It is a matter of taste whether to call quaternions numbers.

Notice that these matrices act on complex 2-vectors $\left[\begin{smallmatrix}x+iy\\u+iv\end{smallmatrix}\right]$, *i.e.,* on a space of four real dimensions. By insisting on real positive determinant we have ruled out the obvious complex matrix to call i, which would be $iI = \left[\begin{smallmatrix}i&0\\0&i\end{smallmatrix}\right]$, with determinant -1. The new i, the matrix $\left[\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right]$ *as a complex matrix*, only behaves and looks like the old one; it is not exactly the old i as embedded in our matrix route to constructing the new and larger system.

Now, the trace (sum of diagonal entries) of the matrix $A = \alpha + i\beta + j\gamma + k\delta$ is just 2α; i, j and k have trace 0, and so do their linear combinations, the **pure quaternions**. For any matrix $\left[\begin{smallmatrix}a&c\\b&d\end{smallmatrix}\right]$ with entries drawn from a system of numbers that commute, like the real or complex ones, we have

$$\begin{aligned}\begin{bmatrix}a&c\\b&d\end{bmatrix}\begin{bmatrix}a&c\\b&d\end{bmatrix} - (a+d)\begin{bmatrix}a&c\\b&d\end{bmatrix} &= \begin{bmatrix}a^2+bc&(a+d)c\\(a+d)b&bc+d^2\end{bmatrix} - \begin{bmatrix}a^2+ad&(a+d)c\\(a+d)b&ad+d^2\end{bmatrix}\\ &= -(ad-bc)\begin{bmatrix}1&0\\0&1\end{bmatrix},\\ \text{or } A^2 &= (\mathsf{trace}\,A)\,A - (\mathsf{det}\,A)\,I\,. \end{aligned}\tag{5.10}$$

So for any matrix with $\mathsf{trace}\,A = 0$ and $\mathsf{det}\,A = 1$, we have $A^2 = -I$. In particular, all the matrices $i\beta + j\gamma + k\delta$ with $\beta^2 + \gamma^2 + \delta^2 = 1$, the unit pure quaternions, have square -1. Instead of just two square roots $\pm i$ for -1, we now have a whole unit sphere of them!

In particular, we have $i^2 = j^2 = k^2 = -1$; with (5.4–5.9), this tells us that the product

$$qr = (\alpha_q + i\beta_q + j\gamma_q + k\delta_q) \times (\alpha_r + i\beta_r + j\gamma_r + k\delta_r)$$

is also a combination $p = \alpha_p + i\beta_p + j\gamma_p + k\delta_p$ of the same four matrices; every term in multiplying out gives us a contribution to the coefficient of 1, i, j, or k (not something else, like $\left[\begin{smallmatrix} i & 0 \\ 0 & i \end{smallmatrix}\right]$, which cannot be written as $\alpha + i\beta + j\gamma + k\delta$). If we write out this multiplication in real and i, j, k components, we can arrange it as

$$\begin{bmatrix} \alpha_p \\ \beta_p \\ \gamma_p \\ \delta_p \end{bmatrix} = \begin{bmatrix} \alpha_q & -\beta_q & -\gamma_q & -\delta_q \\ \beta_q & \alpha_q & -\delta_q & \gamma_q \\ \gamma_q & \delta_q & \alpha_q & -\beta_q \\ \delta_q & -\gamma_q & \beta_q & \alpha_q \end{bmatrix} \begin{bmatrix} \alpha_r \\ \beta_r \\ \gamma_r \\ \delta_r \end{bmatrix} \tag{5.11}$$

$$= \begin{bmatrix} \alpha_r & -\beta_r & -\gamma_r & -\delta_r \\ \beta_r & \alpha_r & \delta_r & -\gamma_r \\ \gamma_r & -\delta_r & \alpha_r & \beta_r \\ \delta_r & \gamma_r & -\beta_r & \alpha_r \end{bmatrix} \begin{bmatrix} \alpha_q \\ \beta_q \\ \gamma_q \\ \delta_q \end{bmatrix} \tag{5.12}$$

and notice that each matrix has all columns of length $\sqrt{\alpha^2 + \beta^2 + \gamma^2 + \delta^2}$ for the appropriate set of $\alpha, \beta, \gamma, \delta$, and all mutually orthogonal (and similarly for rows). Thus the *unit* quaternions—those with determinant $\alpha^2 + \beta^2 + \gamma^2 + \delta^2 = 1$ —act as rotations on the four-real-dimensional $(\alpha, \beta, \gamma, \delta)$-space. But not *all* of the 4D rotations occur this way; for instance,

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is missing.

The fact that quaternions multiply to give quaternions also means that if we multiply two *unit* quaternions q and r, with determinants

$$\mathsf{det}\, q = \alpha_q^2 + \beta_q^2 + \gamma_q^2 + \delta_q^2 = 1\,, \quad \mathsf{det}\, r = \alpha_r^2 + \beta_r^2 + \gamma_r^2 + \delta_r^2 = 1\,,$$

we get another one—a quaternion because it is a product of quaternions, and a unit one because $\mathsf{det}\,(qr) = \mathsf{det}\, q\, \mathsf{det}\, r = 1$, by (1.3). Call the collection[2] of these unit quaternions $\mathcal{Q}$.

Any unit pure quaternion q in $\mathcal{Q}$ has $q^2 = -(\mathsf{det}\, q)1 = -1$, so the combinations $x + qy$ of 1 and q obey the same multiplication rules (among themselves) as $x + iy$. Everything in Section 4 still works in the same way. We get the same kind of parametrized unit circle $\cos t + q \sin t$ for q as for i, and the same compounding behaviour,

$$e^{qt} = \cos t + q \sin t \tag{5.13}$$

[2] As a set of complex matrices, the way we have set it up here, this collection is usually called SU(2), for **S**pecial **U**nitary group in **2** complex variables, but I am skipping the geometry behind 'unitary'. It also arises under the name Spin(3).

by the same arguments. The pure quaternions are the 'infinitesimal members of $\mathcal{Q}$'; they actually correspond to infinitesimal 4D rotations—but not all the infinitesimal 4D rotations—just as $\mathcal{Q}$ corresponds to [some of] the ordinary 4D rotations.

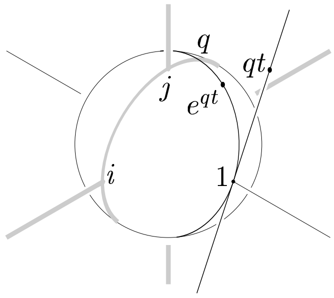

Figure 8. The tangent 3D space to $\mathcal{Q}$ at 1 is a copy of the pure quaternions—the ones shown pale grey here—but based at 1; $qt \mapsto e^{qt}$ wraps a straight line round a great circle in $\mathcal{Q}$. (The picture should have four dimensions, but one's held up in Singapore customs; they all look the same, anyway. See one and you've seen them all. . .)

This picture makes it obvious how to find the inverse of a unit quaternion: just go the other way around, reversing the pure quaternion part. The inverse of e^{qt} is simply

$$e^{-qt} = e^{q(-t)} = \cos t + q\sin(-t) = \cos t - q\sin t\,. \tag{5.14}$$

For a general quaternion we can easily check that

$$\begin{aligned}(\alpha + i\beta + j\gamma + k\delta)(\alpha - i\beta - j\gamma - k\delta) &= \alpha^2 + \beta^2 + \gamma^2 + \delta^2\,, & (5.15)\\ \text{or } p\overline{p} &= \mathsf{det}\,p \quad \text{for short.} & (5.16)\end{aligned}$$

Thus for a unit quaternion $p = \alpha + i\beta + j\gamma + k\delta$ its **conjugate** $\overline{p} = \alpha - i\beta - j\gamma - k\delta$ is also its inverse—the thing we can multiply p by to get 1, or (dynamically) the action which reverses the effect of p when done after p, or which is reversed by p when done before it. For general $p \neq 0$, (5.16) gives immediately $p\left((\mathsf{det}\,p)^{-1}\overline{p}\right) = 1$, so that

$$p^{-1} = \frac{\overline{p}}{\mathsf{det}\,p} = \frac{\alpha - i\beta - j\gamma - k\delta}{\alpha^2 + \beta^2 + \gamma^2 + \delta^2} \tag{5.17}$$

6 3D rotations

The unit quaternions rotate their own 4-space neatly; but we more often want general 3D rotations than special 4D ones. We need one more trick.

If we multiply a quaternion by another on its left, we get the rotation matrix in (5.11); multiplication on the right gives the form (5.12). Left multiplying by q and then q^{-1}, of course, brings us back to where we were, $q^{-1}(qr) = r$ for any r; but because (5.11) and (5.12) are different, left multiplying by q and *right* by q^{-1} leaves us with a net change. We have $qrq^{-1} = r$ if and only if $(qrq^{-1})q = (r)q$, that is, $qr = rq$, that is, q and r commute. But because any real $\alpha 1$ *does* commute with everything, for any q the transformation

$$C_q(r) = qrq^{-1} \tag{6.1}$$

has $C_q(\alpha 1) = \alpha 1$ for all α. Since left multiplication by q^{-1} and right multiplication by q are both rotations, so must C_q be. As C_q leaves the α-axis (the line through 1) fixed, it must also spin the rest of the quaternions around this axis, taking vectors at right angles to 1 to other vectors at right angles to 1. These are just the pure quaternions $i\beta + j\gamma + k\delta$, so

$$r \mapsto qrq^{-1}$$

is a rotation of the 3D space of pure quaternions—whether or not q is a unit quaternion. If it is not, $q = s\widehat{q}$ where $\mathsf{det}\,\widehat{q} = 1$, and because real numbers like s and $1/s$ commute with everything, we have

$$qrq^{-1} = (s\,\widehat{q})\,r(s\widehat{q})^{-1} = (s\,\widehat{q})\,r(\widehat{q})^{-1}\left(\frac{1}{s}\right) = (s)\left(\frac{1}{s}\right)\widehat{q}r\widehat{q}^{-1} = \widehat{q}\,r\,\widehat{q}^{-1}\,, \tag{6.2}$$

which is $C_{\widehat{q}}$ for the unit quaternion $\widehat{q}$. Thus all non-zero quaternions on the same line give the same rotation of 3D pure quaternion space. *All* non-zero quaternions give geometrically exact 3D rotations: it is impossible to drift off, the way one can get lost from the set (4.17–4.22) of rotation matrices. In particular, all real quaternions give the identity 3D rotation, and sign makes no difference;

$$(-q)r(-q)^{-1} \;=\; (-1)qrq^{-1}(-1) \;=\; qrq^{-1}\,.$$

If $\mathsf{det}\,q = 1$, this simplifies by (5.17) to $qr\overline{q}$: otherwise, to $qr\overline{q}/\mathsf{det}\,q$. If we restrict to unit quaternions, any rotation R must be produced by both quaternions $\pm q$ whenever it is produced by any q, if it has any. It turns out that exactly two produce any given R. The least exciting way to see this is to look at the formulæ: for $q = \alpha + i\beta + j\gamma + k\delta$, using the basis $\{i, j, k\}$, the rotation C_q has a matrix given by

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} = \frac{1}{\mathsf{det}\,q}\begin{bmatrix} \alpha^2+\beta^2-\gamma^2-\delta^2 & 2(\beta\gamma-\alpha\delta) & 2(\beta\delta+\alpha\gamma) \\ 2(\beta\gamma+\alpha\delta) & \alpha^2-\beta^2+\gamma^2-\delta^2 & 2(\gamma\delta-\alpha\beta) \\ 2(\beta\delta-\alpha\gamma) & 2(\gamma\delta+\alpha\beta) & \alpha^2-\beta^2-\gamma^2+\delta^2 \end{bmatrix}. \tag{6.3}$$

If $\mathsf{det}\,q = 1$, the multiplier is unneeded (and the diagonal entries simplify a bit).

Conversely, given the rotation matrix, it is easy to show that $[\alpha\ \beta\ \gamma\ \delta]$ must be a real scalar multiple of the unit quaternion

$$\frac{1}{2}\begin{bmatrix} \sqrt{1+m_{11}+m_{22}+m_{33}} \\ \sqrt{1+m_{11}-m_{22}-m_{33}}\,\mathrm{sgn}(m_{32}-m_{23}) \\ \sqrt{1-m_{11}+m_{22}-m_{33}}\,\mathrm{sgn}(m_{13}-m_{31}) \\ \sqrt{1-m_{11}-m_{22}+m_{33}}\,\mathrm{sgn}(m_{21}-m_{12}) \end{bmatrix}^{\text{transpose}}. \tag{6.4}$$

It is more interesting to look at a steady turn. To keep the formulæ simple, let's turn around the i-axis, though in any direction the geometry is the same. Clearly any $C_{\alpha+i\beta}$ has $C_{\alpha+i\beta}(ix) = ix$ for any ix on the i-axis; points on the i-axis do not move, because they commute with $\alpha + i\beta$. So this is indeed the axis of the rotation. Focus on what happens to the j and k components, at right angles to this: for $q = e^{it} = \cos t + i \sin t$, we have

$$
\begin{aligned}
C_{e^{it}}(j\gamma + k\delta) &= (\cos t + i\sin t)(j\gamma + k\delta)(\cos t - i \sin t) \\
&= j\left(\gamma(\cos^2 t - \sin^2 t) - 2\delta \sin t \cos t\right) + k\left(2\gamma \sin t \cos t + \delta(\cos^2 t - \sin^2 t)\right) \\
&= j\tilde{\gamma} + k\tilde{\delta} \\
\text{where } \begin{bmatrix}\tilde{\gamma} \\ \tilde{\delta}\end{bmatrix} &= \begin{bmatrix}\cos 2t & -\sin 2t \\ \sin 2t & \cos 2t\end{bmatrix}\begin{bmatrix}\gamma \\ \delta\end{bmatrix}, \qquad (6.5)
\end{aligned}
$$

so that $j\gamma + k\delta$ has been turned through an angle of $2t$ radians—twice as fast as $\cos t + i \sin t$ is turning (Fig. 9). In particular, any pure quaternion $q = \cos(90°) + q\sin(90°)$ gives a 180° turn around the q-axis, and $q^2 = -1$ gives C_q as the identity. By the time $\cos t + q \sin t$ has reached -1, $C_{\cos t + q \sin t}$ has turned all the way around to the identity 3D rotation. When $\cos t + q \sin t$ gets back to 1, any $C_{\cos t + q\sin t}(p)$ has gone twice around the q-axis.

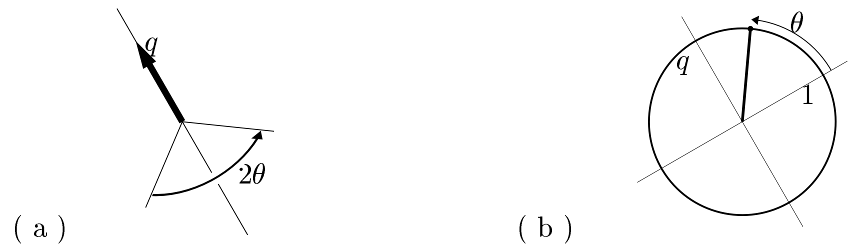

Figure 9. Rotating 3-space by 2θ around the q-axis by a quaternion $\cos\theta + q\sin\theta$ turned only θ from 1 in the plane given by 1 and q. (Fig. 9a is in i, j, k space—which we identify with ordinary 3D coordinate space—while 9b is in the 1, q plane, which meets i, j, k space at right angles.)

This has an important consequence in using (6.4), for a sequence of orientation matirces. Suppose a sensor reports a turn at a steady rate a around the q-axis, with orientation $\mathcal{O}(t)$ rotated through at from the reference position at time t. When $\mathcal{O}(t)$ reaches a 180° turn around the q-axis, the quaternion given by (6.4)—which always chooses $\alpha > 0$, if we take the $\sqrt{\ }$ at the top to mean $\sqrt[+]{\ }$—points along the positive q direction. At the next moments, corresponding to a turn of more than 180°, the continuous choice of next quaternions should obviously have a small *negative* value of α. Instead, by keeping $\alpha > 0$, (6.4) makes the opposite choice, pointing nearly in the $-q$ direction. Blind use of (6.4) would thus give a jump right across the sphere of unit quaternions. This would severely disrupt any filter being used to smooth or predict the sequence! It is therefore necessary to keep checking the dot product $q_n \bullet q_{n+1}$. If that is near 1, fine: if it is nearer -1, switch the signs used for *all* the $(\alpha, \beta, \gamma, \delta)$ in (6.4) until the next negative $q_n \bullet q_{n+1}$ happens. (A similar problem has a similar cure in (x, y, z)-space, with the Ascension Bird 3D sensor, which reports identical numbers for positions (x, y, z) and $(-x, -y, -z)$ in device coordinates. It can be set to 'x always > 0', 'y always > 0', or 'z always > 0'; to detect when the fixed sign should have changed, so all signs should be flipped, check for negative $x_n x_{n+1} + y_n y_{n+1} + z_n z_{n+1}$.)

[Here the speaker shall perform the Soup Plate Trick]

7 Computational aspects

Multiplying two quaternions corresponds to combining the two rotations they give:

$$C_{pq}(r) = (pq)r(pq)^{-1} = pqrq^{-1}p^{-1} = p\left(qrq^{-1}\right)p^{-1} = C_p\left(C_q(r)\right) = \left(C_pC_q\right)(r)\,. \qquad (7.1)$$

Multiplying p and q takes only 16 floating point multiplies, against 27 to multiplying the corresponding 3×3 matrices; saving nearly half the effort. This is very promising numerically, unless the matrix multiplication has been very strongly optimized somewhere in hardware, and quaternion multiplication has not. Taking linear combinations of quaternions (as one must in filtering) takes 4/9 the effort that it does in 9-component rotation matrices. If we are doing interesting things with spatial orientations—always labelled by some way of specifying the rotation necessary to get to the current position from some reference position—this can become important.

Using q directly as a rotation, on the other hand, it takes 24 floating multiplies to compute $C_q(r)$. (Two general quaternion multiplies would need 32, but we know the real part will vanish, so we need only compute the pure quaternion part of the result.) This more than the 9 needed to evaluate a 3×3 matrix on a vector, so actual rotation of a vector—5 particularly, applying the same rotation to many vectors—should most often be done with a matrix. The costs (6.4) and (6.3) of switching from matrices to quaternions (plus a sign check on $q_n \bullet q_{n+1}$ for continuity) and back are significant, so it is only worth making the round trip if substantial work is done at the quaternion end.

When significant work *is* done at the quaternion end, it tends to be easier as well as less costly. The quaternion change between q_n at orientation measurement n and q_{n+1} at the next is just $\overline{q}_n q_{n+1}$; we can compute with these changes as discussed in [1], and estimate the change p from the quaternion describing the current orientation to the the quaternion describing the orientation at time τ ahead of now. Then multiply the current state by e^p, or by some approximation of it. Different approximations give different errors, but accumulated error cannot carry us out of quaternion space, so (6.3) always gives us a true, length-and-right-angle-preserving rotation. (Accumulated round-off error in rotation is nicely illustrated by a small revolving cube in one program in the ISS VR lab; with each Silicon Graphics built in 'rotate( )' matrix multiply that rotates its position a little, it grows imperceptibly in size. Exponentially, given the chance, it would swallow the displayed world in half an hour.)

References

[1] Wayne Lawton, Timothy Poston, and Raghu Raghavan, *Prediction, Smoothing and Control of Rigid Body Trajectories*, in preparation.