ROTATIONS

JOSHUA R. DAVIS

ABSTRACT. We review rotations of two- and three-dimensional space, from basic algebraic manipulation to interpolation and splines (not finished yet). Three different representations of rotations are used: special orthogonal matrices, Euler angles, and quaternions. Methods for converting among these representations are given. The treatment is suitable for anyone who has studied linear algebra.

1. INTRODUCTION

Rigid-body rotations appear in many scientific problems, including physical simulations, computer graphics (drafting, geographical information systems, video games), and geology. These uses ultimately rely on a small set of fundamental mathematical operations with rotations.

- Given some unrotated vectors and some rotated ones, infer the rotation that took the former to the latter.
- Given a rotation T and a vector (or point) $\vec{v}$, compute the rotated vector $T(\vec{v})$.
- Given two rotations T and S, describe the composed rotation $S \circ T$ that corresponds to the combined effect of T followed by S.
- Given a rotation T, invert it, meaning compute the rotation T^{-1} that undoes T.
- Suppose we have a rotation T. Imagine that $\vec{v}$ is the position of some object at time $t = 0$, and that $T(\vec{v})$ is the position at time $t = 1$. What is the position at time $t = 1/2$? It is $S(\vec{v})$, where S is some rotation that is "halfway" between I and T. How do we find S? More generally, can we find a path of rotations from I to T? Can we do it in a canonical (i.e. standard, objective) way?

What is the best way to represent the rotations concretely, so that we may perform such computations with them as simply as possible? We seek simplicity not just because we're lazy, but also because we're *really* lazy — we want computers to do the calculations for us, and the computer programs need to be fast, reliable, and numerically robust. In this paper we try three different descriptions:

- Special orthogonal matrices, meaning matrices M such that $M^\top = M^{-1}$ and $\det M = 1$, are convenient for all kinds of computation, except perhaps interpolating paths.

- Euler angles, which describe how to rotate about three fixed coordinate axes, are not convenient at all for computation, despite their popularity.
- Quaternions, which are like "complex complex numbers", take some getting used to, but they are convenient for all kinds of computation.

All of these systems are theoretically equivalent, in that they describe the same set of rotations. Methods for converting among them are given in this paper. Rotations of two-dimensional space are sufficiently simple that the conversion process is vacuous and no system is significantly better than the others. However, for expository reasons it is desirable to treat two dimensions thoroughly before delving into the complexity of rotations in three dimensions. Along the way some mathematical jargon is introduced, just so that readers who try to learn more from math books will know what to look for.

2. Rotations in General

In this section we define an abstract notion of rotation of $\mathbb{R}^2$ or $\mathbb{R}^3$ (or any $\mathbb{R}^n$, for that matter). This abstract treatment has the benefit of stating the most basic and useful properties of rotations without tying us down to any particular way of writing them. The drawback is that we lose sight of the tactile experience of turning an object around an axis through some angle. But that axis-angle notion of rotation in $\mathbb{R}^3$ is surprisingly difficult to describe mathematically, until you stumble upon the right system. More on that in a moment.

A *rotation* of $\mathbb{R}^n$ is a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ that preserves the dot product and preserves orientation. Explicitly this means that for any vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$ and any scalar $c \in \mathbb{R}$,

(1) $T(\vec{v} + \vec{w}) = T(\vec{v}) + T(\vec{w})$,
(2) $T(c\vec{v}) = cT(\vec{v})$,
(3) $T(\vec{v}) \cdot T(\vec{w}) = \vec{v} \cdot \vec{w}$, and
(4) $\det T > 0$.

The dot product implies notions of angle and length, as follows. The length $|\vec{v}|$ of a vector $\vec{v}$ is given by

$$|\vec{v}| = \sqrt{\vec{v} \cdot \vec{v}}.$$

The angle θ between $\vec{v}$ and $\vec{w}$ satisfies $\vec{v} \cdot \vec{w} = |\vec{v}||\vec{w}| \cos \theta$; in other words,

$$\theta = \arccos \left(\frac{\vec{v} \cdot \vec{w}}{|\vec{v}||\vec{w}|} \right).$$

Because any rotation T preserves the dot product, it follows that T preserves lengths and angles:

$$\begin{aligned} |T(\vec{v})| &= |\vec{v}|, \\ \arccos \left(\frac{T(\vec{v}) \cdot T(\vec{w})}{|T(\vec{v})||T(\vec{w})|} \right) &= \arccos \left(\frac{\vec{v} \cdot \vec{w}}{|\vec{v}||\vec{w}|} \right). \end{aligned}$$

One can also deduce that a rotation preserves area/volume/etc. In terms of linear algebra, this means that its determinant is ± 1. But we require a rotation to have positive determinant, so its determinant must be 1. Linear transformations with negative determinant involve reversals of orientation (reflections, or flips of space), and we don't want those. In short, the notion of rotation we've laid out here transforms space without distorting or flipping it in any way.

If T and S are two rotations, then the combined effect of T followed by S is the composition of functions $S \circ T$. We have to write it in this order because the vector being rotated is written on the right:

$$(S \circ T)(\vec{v}) = S(T(\vec{v})).$$

Since both T and S are linear, so is $S \circ T$; since both preserve the dot product, so does $S \circ T$; since both have positive determinant, so does $S \circ T$. This shows that the composition of two rotations is a rotation.

Pause for a moment to consider that fact: The combined effect of a rotation T followed by another rotation S is itself a rotation. This is not at all obvious, if one thinks of rotations in terms of axes and angles. If you rotate $\mathbb{R}^3$ about one axis, and then rotate it again about some other axis, then the net effect is a rotation about some third axis? Really? Which one? How much have you rotated about it?

Before we get to any concrete description of rotations, let's finish laying out their basic properties in the abstract. The set of all rotations of $\mathbb{R}^n$ forms a *group* under composition, which simply means that the following four properties are satisfied.

- Closure: If T and S are rotations, then so is $S \circ T$, as we've already seen.
- Associativity: For any rotations T, S, and U, $U \circ (S \circ T) = (U \circ S) \circ T$. This property is always enjoyed by compositions of functions; it just says that
$$(U \circ (S \circ T))(\vec{v}) = U(S(T(\vec{v}))) = ((U \circ S) \circ T)(\vec{v}).$$
- Identity: The identity I — the linear transformation that does nothing, as in $I\vec{v} = \vec{v}$ — is trivially a rotation. For any rotation T, $I \circ T = T \circ I = T$.
- Inverses: Any rotation T is invertible, and its inverse T^{-1} is also a rotation. The inverse satisfies $T \circ T^{-1} = T^{-1} \circ T = I$.

The rotations of $\mathbb{R}^2$ commute with each other, meaning $S \circ T = T \circ S$. However, the rotations of $\mathbb{R}^3$ (and higher $\mathbb{R}^n$) do not commute. Order matters. This is especially significant when inverting compositions: The inverse of $S \circ T$ is $T^{-1} \circ S^{-1}$, not $S^{-1} \circ T^{-1}$.

We need to establish two conventions. First, a vector in $\mathbb{R}^n$ is an $n \times 1$ column matrix. Second, we always measure angles in radians. Radians enjoy a number of inherent mathematical properties that other systems of angle measurement (degrees, gradians, etc.) do not. For this reason, most programming languages (Excel, MATLAB, Mathematica, C, etc.) measure angles in radians by default. To convert from degrees to radians, multiply by $2\pi/360^\circ$. For example $360^\circ = 2\pi$, $180^\circ = \pi$, and $90^\circ = \pi/2$. To convert from radians to degrees, divide by $2\pi/360^\circ$. If you like, just view the symbol $^\circ$ as the unitless constant $2\pi/360 \approx 0.0174533$.

3. Rotations as Special Orthogonal Matrices

Rotations are linear transformations, which are the province of linear algebra. In linear algebra, one typically represents linear transformations as matrices, so that will be our first approach to describing rotations.

An $n \times n$ matrix M with real entries is said to be *orthogonal* if its inverse equals its transpose: $M^{-1} = M^\top$, or, in other words,

$$MM^\top = M^\top M = I.$$

Transposes are relevant to us because the dot product of any two vectors $\vec{v}$ and $\vec{w}$ can be written in terms of matrix multiplication like this:

$$\vec{v} \cdot \vec{w} = \vec{v}^\top \vec{w}.$$

If we multiply $\vec{v}$ and $\vec{w}$ by an orthogonal matrix M, then

$$(M\vec{v}) \cdot (M\vec{w}) = (M\vec{v})^\top M\vec{w} = \vec{v}^\top M^\top M\vec{w} = \vec{v}^\top I\vec{w} = \vec{v}^\top \vec{w} = \vec{v} \cdot \vec{w}.$$

That is, multiplication by an orthogonal matrix preserves the dot product. Thus it also preserves lengths, angles, and area/volume/etc. Its determinant is ± 1. If it has determinant 1 then it's called a *special orthogonal* matrix and it represents a rotation of $\mathbb{R}^n$. Orthogonal matrices with determinant -1 represent rotations with flips, which we want to avoid.

But what does a special orthogonal matrix look like? Well, it turns out that a matrix is orthogonal if and only if its columns form an *orthonormal basis* of $\mathbb{R}^n$, meaning that each column, regarded as a vector, has length 1 and is perpendicular to the others. Equivalently, a matrix is orthogonal if and only if its rows form an orthonormal basis. It follows that every entry in an orthogonal matrix is between -1 and 1, inclusively. Here are two examples of orthogonal matrices:

$$\begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}, \quad \begin{bmatrix} 1/2 & -\sqrt{3}/2 & 0 \\ 3/4 & \sqrt{3}/4 & -1/2 \\ \sqrt{3}/4 & 1/4 & \sqrt{3}/2 \end{bmatrix}.$$

Both have determinant 1, so they're special orthogonal. The numbers in them look a little uniform because I constructed them using some convenient Euler angles. More on that in the next two sections.

To apply the rotation T represented by the special orthogonal matrix M to a vector $\vec{v}$, we simply multiply them as matrices:

$$T(\vec{v}) = M\vec{v}.$$

To compose two rotations T and S, we multiply their matrices. To be precise, if T is represented by the matrix M and S is represented by the matrix N, then the rotation $S \circ T$ comprising T followed by S is represented by the product matrix NM:

$$(S \circ T)(\vec{v}) = S(T(\vec{v})) = NM\vec{v}.$$

Inversion of rotations corresponds to inversion of matrices, meaning that if M represents T, then M^{-1} represents T^{-1}. In general, inversion of matrices is neither fast nor numerically robust, but inversion of orthogonal matrices is both, because it is just transposition:

$$M^{-1} = M^\top.$$

We'll leave the question of interpolation until later. For all other basic operations, special orthogonal matrices are quite computationally convenient. But we still don't know how they relate to the idea of turning an object about an axis through some angle. That comes next.

4. Two-Dimensional Rotations as Described by Angles

In this section we describe rotations of the plane $\mathbb{R}^2$ by specifying the angles through which they rotate. By convention, we always measure angles counterclockwise about the origin; that is, we follow a right-hand rule about an axis pointing out of the plane toward the viewer. Remember that we always measure angles in radians.

Given an angle θ and a two-dimensional vector $\vec{v}$ representing a point in the plane, how do we rotate $\vec{v}$ counterclockwise about the origin by θ? Perhaps the simplest way is to write down a special orthogonal matrix M_θ that represents rotation by θ, and then to multiply M_θ by $\vec{v}$. The matrix is

$$M_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Using the fundamental trigonometric identity $\sin^2\theta + \cos^2\theta = 1$, it is easy to see that M_θ is special orthogonal for any θ. Namely,

$$\begin{aligned} M_\theta M_\theta^\top &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos^2\theta + \sin^2\theta & \cos\theta\sin\theta - \sin\theta\cos\theta \\ \sin\theta\cos\theta - \cos\theta\sin\theta & \sin^2\theta + \cos^2\theta \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

(One should also check that $M_\theta^\top M_\theta = I$ and that $\det M_\theta = 1$.) For example, if $\theta = \pi/6 = 30°$, then

$$M_{\pi/6} = \begin{bmatrix} \cos\pi/6 & -\sin\pi/6 \\ \sin\pi/6 & \cos\pi/6 \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \approx \begin{bmatrix} 0.87 & -0.5 \\ 0.5 & 0.87 \end{bmatrix}.$$

If we multiply this by $\vec{v} = (1, 0)$, then sure enough we get $\vec{v}$ rotated by $\pi/6$:

$$\begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix} \approx \begin{bmatrix} 0.87 \\ 0.5 \end{bmatrix}.$$

Conversely, every special orthogonal matrix

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

is equal to M_θ for some angle θ. Namely, if $m_{21} \geq 0$, then $\theta = \arccos m_{11}$; if $m_{21} < 0$, then $\theta = -\arccos m_{11}$. Using the fact the columns of M are orthonormal, one can then check that $m_{12} = -\sin\theta$, $m_{21} = \sin\theta$, and $m_{22} = \cos\theta$, as desired.

If we rotate two-dimensional space through some angle θ and then through some angle ϕ, then the combined effect is to rotate through the angle $\phi + \theta$. So composing two rotations amounts to adding their angles. Of course, composing the rotations also corresponds to multiplying their matrices, so it had better be true that $M_\phi M_\theta = M_{\phi+\theta}$. We verify this now:

$$\begin{aligned} M_\phi M_\theta &= \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos\phi\cos\theta - \sin\phi\sin\theta & -\cos\phi\sin\theta - \sin\phi\cos\theta \\ \sin\phi\cos\theta + \cos\phi\sin\theta & -\sin\phi\sin\theta + \cos\phi\cos\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos(\phi+\theta) & -\sin(\phi+\theta) \\ \sin(\phi+\theta) & \cos(\phi+\theta) \end{bmatrix} \\ &= M_{\phi+\theta}. \end{aligned}$$

Here we have used the trigonometric identities

$$\begin{aligned} \cos(\phi+\theta) &= \cos\phi\cos\theta - \sin\phi\sin\theta, \\ \sin(\phi+\theta) &= \sin\phi\cos\theta + \cos\phi\sin\theta. \end{aligned}$$

Similarly, the inverse of the rotation through θ is the rotation through $-\theta$. Using the identities $\cos(-\theta) = \cos\theta$ and $\sin(-\theta) = -\sin\theta$, you can check that $M_{-\theta} = M^\top = M_\theta^{-1}$.

Philosophically, all facts about rotations in $\mathbb{R}^2$ boil down to trig identities, and all important trig identities come up in the proofs of these statements, because *trigonometry is the study of the rotations of* $\mathbb{R}^2$. This rotation matrix stuff is simply a rephrasing of trigonometry.

5. Three-Dimensional Rotations as Described by Angles

In three dimensions, angular representations of rotations become much more complicated. Let us again agree that all angles are given in radians and that they describe rotations counterclockwise about their axes according to a right-hand rule. For example, suppose the desired rotation axis is the z-axis. Hold your right hand in a fist, with your thumb pointing toward your face. Your thumb represents the positive z-axis, and your fingers are curling in a counterclockwise direction around that axis. A rotation of $\pi/2$ rotates everything 90° counterclockwise, in the direction that your fingers are pointing. For example, the point $(0, 1, 0)$ rotates to $(-1, 0, 0)$. If a clockwise or left-handed convention were in effect, then $(0, 1, 0)$ would go to $(1, 0, 0)$.

With these conventions, it turns out that a rotation about the z-axis through an angle γ corresponds to this 3×3 matrix Z_γ:

$$Z_\gamma = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

It is easy to check that Z_γ is special orthogonal for any γ, using trig identities as in the preceding section. Similarly, a rotation about the y-axis through an angle β is given by

$$Y_\beta = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix},$$

and a rotation about the x-axis through an angle α is given by

$$X_\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}.$$

Notice that the minus sign appears in a "different place" in the rotation about the y-axis.

Now imagine a rotation about the x-axis by some angle α, followed by a rotation about the y-axis by an angle β, followed by a rotation about the z-axis by an angle γ. For short, we'll call this sort of composition an *x-y-z-rotation*. It corresponds to the product of matrices

$$\begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}.$$

As always, the matrices are written in "reverse order" because we apply them by putting a vector on the right. The product of these matrices is

$$\begin{bmatrix} \cos\gamma\cos\beta & \cos\gamma\sin\beta\sin\alpha - \sin\gamma\cos\alpha & \cos\gamma\sin\beta\cos\alpha + \sin\gamma\sin\alpha \\ \sin\gamma\cos\beta & \sin\gamma\sin\beta\sin\alpha + \cos\gamma\cos\alpha & \sin\gamma\sin\beta\cos\alpha - \cos\gamma\sin\alpha \\ -\sin\beta & \cos\beta\sin\alpha & \cos\beta\cos\alpha \end{bmatrix}.$$

The three angles α, β, and γ are called the *Euler angles* of the rotation. It turns out that any rotation of three-dimensional space is completely described by its three Euler angles. In other words, we have three degrees of freedom in choosing how to rotate space. To see this intuitively, think of how we might orient a coordinate frame described by perpendicular unit vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$. First we might choose where the vector $\vec{a}$ goes. It can rotate in any direction, so that its tip is anywhere on the unit sphere. Once we have oriented $\vec{a}$, we choose where $\vec{b}$ goes. But $\vec{b}$ must be perpendicular to $\vec{a}$, and that forces it to lie along a circle. Once $\vec{a}$ and $\vec{b}$ are chosen, the orientation of $\vec{c}$ is determined, because it must be perpendicular to $\vec{a}$ and

$\vec{b}$ according to the right-hand rule. So we have two degrees of freedom in choosing where $\vec{a}$ goes, one degree of freedom in choosing where $\vec{b}$ goes, and no freedom in choosing where $\vec{c}$ goes. So we have three degrees of freedom in all.

Now if

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$$

is an arbitrary special orthogonal matrix, how do we find its Euler angles α, β, and γ? We compare M to the complicated x-y-z-rotation matrix given above and solve for α, β, and γ entry by entry. For starters, m_{31} must equal $-\sin\beta$, so either $\beta = -\arcsin m_{31}$ or $\beta = \pi + \arcsin m_{31}$. Pick one and compute $\cos\beta$. Now it must be true that $\cos\alpha = m_{33}/\cos\beta$ and $\sin\alpha = m_{32}/\cos\beta$. As in the previous section, infer α from its sine and cosine. Use a similar process to infer γ from m_{11} and m_{21}. At this point we have values for all three angles. Compute some of the other entries in M; if our α, β, and γ do not produce the correct results, then repeat the process, using the other value for β. (Of course, if you're programming this you'll want to be more systematic, but then I suggest you find some code on the Internet.)

This procedure breaks down in the important special case when $m_{31} = \pm 1$. For then $\beta = \mp\pi/2$ and $\cos\beta = 0$, so we cannot divide by $\cos\beta$. In this special case the x-y-z-rotation matrix simplifies to

$$\begin{bmatrix} 0 & -\sin(\gamma \pm \alpha) & \mp\cos(\gamma \pm \alpha) \\ 0 & \cos(\gamma \pm \alpha) & \mp\sin(\gamma \pm \alpha) \\ \pm 1 & 0 & 0 \end{bmatrix}.$$

That is, when $\beta = -\pi/2$ the rotation depends only on the single quantity $\gamma + \alpha$, rather than on γ and α separately. Similarly, when $\beta = \pi/2$ the rotation depends only on $\gamma - \alpha$.

Usually when we specify a value for one of the three Euler angles we expect to have two degrees of freedom remaining — we still get to choose γ and α, right? But specifying that $\beta = \mp\pi/2$ leaves us with only one degree of freedom, in that only $\gamma \pm \alpha$ matters. This unexpected loss of a degree of freedom is called *gimbal lock*. It is loosely analogous to the fact that at the north and south poles of the Earth, where the latitude is $\pm\pi/2$, the longitude is undefined.

Gimbal lock is a significant defect of the Euler angle description of rotations. Because of it, the $\cos\beta = 0$ case must be treated specially in computations; furthermore, when $\cos\beta$ is nonzero but close to zero, numerical procedures to determine α and γ are imprecise. We would prefer a description of rotations in which we always have three well-behaved degrees of freedom.

It should be noted that some authors do the rotations with respect to different axes or in a different order; for example, [!!Akenine] uses what I would call y-x-z-rotations. Such inconsistencies are inconvenient but not a theoretical difficulty. Just about any convention works as well as any other. As long as we agree on some convention, we can completely describe any rotation by specifying three Euler angles.

Unfortunately, just about every kind of computation is inconvenient when expressed in Euler angles. The composition of two x-y-z-rotations is something that might be called an x-y-z-x-y-z-rotation; how do we find its Euler angles, to reexpress it as a single x-y-z-rotation? No, you cannot just add the angles about the x-, y-, and z-axes. The inverse of an x-y-z-rotation is a z-y-x-rotation. How do we express that as an x-y-z-rotation? You cannot just negate the angles. Probably the most practical way to compose and invert rotations described by Euler angles is to convert over to special orthogonal matrices, multiply or invert, and then infer the Euler angles from the resulting matrix M using the procedure described above. However, this procedure is complicated, slow, and numerically imprecise. In Euler angles it is also difficult to specify rotations about arbitrary axes and to interpolate between two rotations. So we need a better system.

6. Two-Dimensional Rotations as the Unit Complex Numbers

In this section we equate the group of rotations of $\mathbb{R}^2$ with the group of unit complex numbers. This is only superficially different from the angular model presented earlier, and it offers no practical advantage. However, the reader who understands this presentation will be better prepared to understand the three-dimensional rotations as the unit quaternions in the next section.

Recall that a *complex number* is a number of the form $a + ib$, where a and b are real numbers and i is a fixed *imaginary unit*, a number such that $i^2 = -1$. Therefore the product of two complex numbers is

$$(a + ib)(c + id) = ac + aid + ibc + ibid = (ac - bd) + i(ad + bc).$$

We identify the set of complex numbers with the real plane by identifying $a+ib$ with the point (a, b). The *norm* $|a+ib|$ of $a+ib$ is simply its distance $\sqrt{a^2 + b^2}$ from the origin. A *unit complex number* is a complex number with norm 1; it lies on the unit circle in the plane. If $|a + ib| = 1$, then there is an angle θ such that $a = \cos\theta$ and $b = \sin\theta$. So the unit complex numbers are exactly those of the form $\cos\theta + i\sin\theta$ for some angle θ. We can also express arbitrary complex numbers $a + ib$ in polar form

$$r\cos\theta + ir\sin\theta = r(\cos\theta + i\sin\theta),$$

where $r = |a + ib|$ and θ is the angle corresponding to the unit complex number $(a + ib)/r$. The angle is undefined when $r = 0$.

By inspecting the Taylor series for e^z, $\cos z$, and $\sin z$, it is easy to discover Euler's formula

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

(The legendary equation $e^{i\pi} + 1 = 0$ follows.) So the complex number of norm r and angle θ can be written succinctly as

$$r(\cos\theta + i\sin\theta) = re^{i\theta}.$$

Multiplication of complex numbers is most naturally done in this form:

$$se^{i\phi}re^{i\theta} = sre^{i(\phi+\theta)}.$$

For any vector $\vec{v} \in \mathbb{R}^2$, we can multiply $\vec{v}$ by any complex number $re^{\imath\theta}$ by regarding $\vec{v}$ as a complex number, doing complex multiplication, and regarding the resulting number $re^{i\theta}\vec{v}$ as a vector in $\mathbb{R}^2$ again. For any complex number $re^{i\theta}$, the resulting map $\mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation.

Under this view of complex numbers as linear transformations of $\mathbb{R}^2$, the unit complex number $e^{i\alpha}$ represents a counterclockwise rotation of the two-dimensional plane through the angle α. To see this, write $\vec{v}$ as $\vec{v} = |\vec{v}|e^{i\theta}$, where θ is the heading of $\vec{v}$, measured counterclockwise from the x-axis. Then

$$e^{i\alpha}|\vec{v}|e^{i\theta} = |\vec{v}|e^{i(\alpha+\theta)}.$$

The length of the rotated vector manifestly equals the length $|\vec{v}|$ of the original, but the heading of the vector has rotated from θ to $\alpha + \theta$.

To compose two rotations $e^{i\alpha}$ and $e^{i\beta}$, we simply multiply them. The composition is $e^{i\beta} \cdot e^{i\alpha} = e^{i(\beta+\alpha)}$. The identity is $e^0 = 1$. The inverse of $e^{i\alpha}$ is $e^{-i\alpha}$, since $e^{i\alpha} \cdot e^{-i\alpha} = e^0 = 1$. These agree with the composition and inversion operations in the angular description of rotations of $\mathbb{R}^2$.

If $a + ib = e^{i\alpha}$ is a unit complex number describing a rotation through α, then the corresponding special orthogonal matrix is

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

7. Quaternions

It turns out that the complex numbers can be extended to an even larger number system, the quaternions, that model rotations of three-dimensional space excellently. The development of the concept takes a little while, but it's worth it. Quaternions make all sorts of rotation operations — application to vectors, composition, inversion, interpolation, rotation about arbitrary axes — easy, fast, and numerically robust.

Recall that a complex number is one of the form $a + ib$, where a and b are real numbers and i is a fixed imaginary unit satisfying $i^2 = -1$. Analogously, a *quaternion* is a number of the form $a + ib_1 + jb_2 + kb_3$, where a, b_1, b_2, and b_3 are real numbers and i, j, and k are distinct fixed *imaginary units* satisfying $i^2 = j^2 = k^2 = -1$. We must also specify how i, j, and k interact, and this is where things become strange:

$$ij = k, \quad jk = i, \quad ki = j.$$

It follows that $ijk = -1$ and that the imaginary units anticommute with each other:

$$ij = -ji, \quad jk = -kj, \quad ki = -ik.$$

So while the complex numbers are commutative, the quaternions are not. In the end this is appropriate, since the two-dimensional rotations commute and the three-dimensional rotations do not. The quaternions do enjoy all of the other basic algebraic properties that we like, such as associativity.

As an aside, it is worth pointing out that

$$a + ib_1 + jb_2 + kb_3 = (a + ib_1) + (b_2 + ib_3)j.$$

The quaternions are "complex complex numbers", in that they are complex numbers $A + Bj$ with coefficients $A = a + ib_1$, $B = b_2 + ib_3$ that are themselves complex. The quaternions for which $b_2 = b_3 = 0$ are ordinary complex numbers; the ones for which $b_1 = 0$ as well are the real numbers. The quaternions extend the complex number system, similarly to how the complex numbers extend the real number system. This is essentially the point of view that Rowan Hamilton held when he invented them in the mid-1800s.

Just as it is customary to identify the complex number $a + ib$ with the two-dimensional vector (a, b), it is customary to identify $a + ib_1 + jb_2 + kb_3$ with the four-dimensional vector (a, b_1, b_2, b_3). But now the notation takes a strange turn. It is common to write

$$a + ib_1 + jb_2 + kb_3 = a + (b_1, b_2, b_3) = a + \vec{b},$$

recording the three imaginary coordinates together in a three-dimensional vector $\vec{b} = (b_1, b_2, b_3)$. This is an abusive notation, and care must be taken in using it. The expression $a + \vec{b}$ does *not* represent the sum of the scalar a and the vector $\vec{b}$; that wouldn't make any sense. It just represents a quaternion with real part a and imaginary parts given by the three entries in $\vec{b}$.

Why put up with this bizarre notation? Well, using it we can derive an interesting and useful formula for quaternion multiplication:

$$(a + \vec{b})(c + \vec{d}) = (ac - \vec{b} \cdot \vec{d}) + (\vec{b} \times \vec{d} + \vec{d}a + \vec{b}c),$$

where $\cdot$ and $\times$ denote the usual dot and cross products of three-dimensional vectors! Keep in mind that the first term on the right is a scalar — the real part of the quaternion — while the second part is a vector representing the imaginary part. For example, when the real parts a and c are 0, then we obtain the formula

$$(0 + \vec{b})(0 + \vec{d}) = -b \cdot d + \vec{b} \times \vec{d}.$$

By this point, your interest should be piqued.

The real number 1 is the multiplicative identity in quaternions, as you'd expect: $(a + \vec{b})1 = a + \vec{b} = 1(a + \vec{b})$. You can see this by writing 1 as $1 + \vec{0}$ and then using the dot/cross multiplication formula:

$$(a + \vec{b})(1 + \vec{0}) = (a - \vec{b} \cdot \vec{0}) + (\vec{b} \times \vec{0} + \vec{0}a + \vec{b}1) = a + \vec{b},$$

and similarly $(1+\vec{0})(a+\vec{b}) = a+\vec{b}$.

The *norm* of a quaternion $a + ib_1 + jb_2 + kb_3$ is

$$|a + ib_1 + jb_2 + kb_3| = \sqrt{a^2 + b_1^2 + b_2^2 + b_3^2} = \sqrt{a^2 + \vec{b}\cdot\vec{b}} = \sqrt{a^2 + |\vec{b}|^2}.$$

The *unit* quaternions are those with norm 1. The product of any two unit quaternions is again a unit quaternion. The real numbers $1 = 1 + \vec{0}$ and $-1 = -1 + \vec{0}$ are both unit quaternions; they are the only unit quaternions with zero imaginary part. If $a + \vec{b}$ is unit, then

$$(a+\vec{b})(a-\vec{b}) = (a^2 + \vec{b}\cdot\vec{b}) + (\vec{b}\times\vec{b} + \vec{b}a - \vec{b}a) = 1 + \vec{0}.$$

So the inverse of any unit quaternion $a + \vec{b}$ is simply $a - \vec{b}$. Explicitly,

$$(a + ib_1 + jb_2 + kb_3)^{-1} = a - ib_1 - jb_2 - kb_3.$$

So far we have discussed quaternions in their rectangular form. It is also profitable, in analogy with the complex numbers, to consider a polar form. If we begin with a unit vector $\vec{u} \in \mathbb{R}^3$ and an angle α, then we can construct a quaternion with real part $\cos\alpha$ and imaginary part $\vec{u}\sin\alpha$. Behold:

$$\begin{aligned}
|\cos\alpha + \vec{u}\sin\alpha| &= \sqrt{\cos^2\alpha + u_1^2\sin^2\alpha + u_2^2\sin^2\alpha + u_3^2\sin^2\alpha} \\
&= \sqrt{\cos^2\alpha + \sin^2\alpha|\vec{u}|^2} \\
&= \sqrt{\cos^2\alpha + \sin^2\alpha} \\
&= 1.
\end{aligned}$$

So $\cos\alpha + \vec{u}\sin\alpha$ is a unit quaternion. Conversely, all unit quaternions

$$a + \vec{b} = a + ib_1 + jb_2 + kb_3$$

are of this form. To find $\vec{u}$ and α from a and $\vec{b}$, use this procedure: If $|\vec{b}| = \sqrt{b_1^2 + b_2^2 + b_3^2} = 0$, then $\alpha = 0$ and it doesn't matter what $\vec{u}$ is; in the typical case, when $|\vec{b}| \neq 0$, let $\vec{u} = \vec{b}/|\vec{b}|$ and $\alpha = \arccos a$. Then $\cos\alpha + \vec{u}\sin\alpha = a + \vec{b}$. So the unit quaternions are exactly those of the form $\cos\alpha + \vec{u}\sin\alpha$ for $\vec{u}$ a unit vector. This is analogous to the fact that the unit complex numbers are exactly those of the form $\cos\alpha + i\sin\alpha$.

Negation of α, $\vec{u}$, and $\cos\alpha + \vec{u}\sin\alpha$ produce interesting effects. First, negating α is tantamount to negating $\vec{u}$, which is tantamount to inverting the quaternion:

$$\cos(-\alpha) + \vec{u}\sin(-\alpha) = \cos\alpha - \vec{u}\sin\alpha = (\cos\alpha + \vec{u}\sin\alpha)^{-1}.$$

Negating *both* α and $\vec{u}$ leaves the quaternion unchanged:

$$\cos(-\alpha) + (-\vec{u})\sin(-\alpha) = \cos\alpha + (-\vec{u})(-\sin\alpha) = \cos\alpha + \vec{u}\sin\alpha.$$

The procedure given above for converting from rectangular to polar form always produces angles $\alpha = \arccos a$ between 0 and π, which may seem biased, but now we see that we could express the same quaternion using $\alpha =$

$-\arccos a$ and $\vec{u} = -\vec{b}/|\vec{b}|$, so there is no bias. Finally, negating $\cos\alpha + \vec{u}\sin\alpha$ itself produces

$$-(\cos\alpha + \vec{u}\sin\alpha) = (-\cos\alpha) + \vec{u}(-\sin\alpha) = \cos(\alpha+\pi) + \vec{u}\sin(\alpha+\pi),$$

which is also a unit quaternion, with angle $\alpha + \pi$. This becomes highly significant in the next section.

To multiply two unit quaternions $\cos\alpha + \vec{u}\sin\alpha$ and $\cos\beta + \vec{v}\sin\beta$, we can use the dot/cross formula:

$$\begin{aligned} & (\cos\beta + \vec{v}\sin\beta)(\cos\alpha + \vec{u}\sin\alpha) \\ = \; & (\cos\beta\cos\alpha - \vec{v}\cdot\vec{u}\sin\beta\sin\alpha) \\ & + (\vec{v}\times\vec{u}\sin\beta\sin\alpha + \vec{u}\sin\alpha\cos\beta + \vec{v}\sin\beta\cos\alpha). \end{aligned}$$

This isn't especially pretty, except in the special case when $\vec{v} = \vec{u}$:

$$\begin{aligned} & (\cos\beta + \vec{u}\sin\beta)(\cos\alpha + \vec{u}\sin\alpha) \\ = \; & (\cos\beta\cos\alpha - \vec{u}\cdot\vec{u}\sin\beta\sin\alpha) \\ & + (\vec{u}\times\vec{u}\sin\beta\sin\alpha + \vec{u}\sin\alpha\cos\beta + \vec{u}\sin\beta\cos\alpha) \\ = \; & (\cos\beta\cos\alpha - \sin\beta\sin\alpha) + \vec{u}(\sin\alpha\cos\beta + \sin\beta\cos\alpha) \\ = \; & \cos(\beta+\alpha) + \vec{u}\sin(\beta+\alpha). \end{aligned}$$

So when two unit quaternions are built from the same unit vector $\vec{u}$, multiplying them amounts to adding their angles.

By analogy with the formula $e^{i\alpha} = \cos\alpha + i\sin\alpha$ for unit complex numbers, we can define an exponential notation for unit quaternions:

$$e^{\vec{u}\alpha} = \cos\alpha + \vec{u}\sin\alpha.$$

Again, this notation is abusive. It does *not* represent the number e raised to the vector $\vec{u}\alpha$; it is simply shorthand for $\cos\alpha + \vec{u}\sin\alpha$, which in turn is shorthand for the rectangular form $\cos\alpha + iu_1\sin\alpha + ju_2\sin\alpha + ku_3\sin\alpha$. But it ends up having some nice properties. We've already proved

$$\begin{aligned} e^{\vec{u}\beta}e^{\vec{u}\alpha} &= e^{\vec{u}(\beta+\alpha)}, \\ e^{(-\vec{u})\alpha} &= e^{\vec{u}(-\alpha)}, \\ e^{(-\vec{u})(-\alpha)} &= e^{\vec{u}\alpha}, \\ -e^{\vec{u}\alpha} &= e^{\vec{u}(\alpha+\pi)}. \end{aligned}$$

Since $e^{(-\vec{u})\alpha} = e^{\vec{u}(-\alpha)}$, we can write $e^{-\vec{u}\alpha}$ to mean either of these, without ambiguity. We know that $(e^{\vec{u}\alpha})^{-1} = e^{-\vec{u}\alpha}$. More generally, $e^{\vec{u}\alpha}$ can be raised to any real exponent t by the formula

$$(e^{\vec{u}\alpha})^t = e^{\vec{u}\alpha t}.$$

So the exponential notation enjoys many of the algebraic properties that one would expect. However, it is *not* true that

$$e^{\vec{v}\beta}e^{\vec{u}\alpha} = e^{\vec{v}\beta+\vec{u}\alpha}$$

when $\vec{u}$ and $\vec{v}$ are distinct vectors. It's not even clear what the right-hand side would mean. This sort of multiplication has to be carried out in rectangular form:

$$\begin{aligned} e^{\vec{v}\beta}e^{\vec{u}\alpha} &= (\cos\beta + \vec{v}\sin\beta)(\cos\alpha + \vec{u}\sin\alpha) \\ &= (\cos\beta\cos\alpha - \vec{v}\cdot\vec{u}\sin\beta\sin\alpha) \\ &\quad + (\vec{v}\times\vec{u}\sin\beta\sin\alpha + \vec{u}\sin\alpha\cos\beta + \vec{v}\sin\beta\cos\alpha). \end{aligned}$$

8. Three-Dimensional Rotations as Unit Quaternions

The point of all this is that the unit quaternion $e^{\vec{u}\alpha} = \cos\alpha + \vec{u}\sin\alpha$ represents a rotation of $\mathbb{R}^3$ about the axis $\vec{u}$ through an angle of 2α, according to the right-hand rule. For example, if $\vec{u} = (1, 0, 0)$ and $\alpha = \pi/4$, then $e^{\vec{u}\alpha}$ represents a rotation about the x-axis through $\pi/2$.

Here's how this works. Given a unit quaternion $a + \vec{b}$ and a three-dimensional vector $\vec{v}$ to rotate, we first regard $\vec{v}$ as the quaternion $0 + \vec{v}$. Then we *conjugate* $0 + \vec{v}$ by $a + \vec{b}$ — meaning, we compute

$$(a + \vec{b})(0 + \vec{v})(a + \vec{b})^{-1}.$$

It turns out that the answer's real part is 0, and that its imaginary part is the rotated vector we wanted. In detail,

$$\begin{aligned} & (a + \vec{b})(0 + \vec{v})(a + \vec{b})^{-1} \\ = \;& (a + \vec{b})(0 + \vec{v})(a - \vec{b}) \\ = \;& \Big((0 - \vec{b}\cdot\vec{v}) + (\vec{b}\times\vec{v} + \vec{v}a)\Big)(a - \vec{b}) \\ = \;& \Big(-a\vec{b}\cdot\vec{v} + (\vec{b}\times\vec{v} + \vec{v}a)\cdot\vec{b}\Big) \\ & + \Big(-(\vec{b}\times\vec{v} + \vec{v}a)\times\vec{b} + \vec{b}(\vec{b}\cdot\vec{v}) + (\vec{b}\times\vec{v} + \vec{v}a)a\Big) \\ = \;& 0 + \Big(-(\vec{b}\times\vec{v})\times\vec{b} - \vec{v}\times\vec{b}a + \vec{b}(\vec{b}\cdot\vec{v}) + \vec{b}\times\vec{v}a + \vec{v}a^2\Big) \\ = \;& 0 + \Big(-(\vec{b}\times\vec{v})\times\vec{b} + (\vec{b}\times\vec{v})2a + \vec{b}(\vec{b}\cdot\vec{v}) + \vec{v}a^2\Big) \\ = \;& 0 + \Big(-2(\vec{b}\times\vec{v})\times\vec{b} + (\vec{b}\times\vec{v})2a + \vec{v}(\vec{b}\cdot\vec{b} + a^2)\Big) \\ = \;& 0 + \Big(-2(\vec{b}\times\vec{v})\times\vec{b} + (\vec{b}\times\vec{v})2a + \vec{v}\Big). \end{aligned}$$

Here we have used identities such as

$$\begin{aligned} (\vec{b}\times\vec{v})\cdot\vec{b} &= 0, \\ -\vec{v}\times\vec{b} &= \vec{b}\times\vec{v}, \\ -(\vec{b}\times\vec{v})\times\vec{b} &= \vec{b}(\vec{b}\cdot\vec{v}) - \vec{v}(\vec{b}\cdot\vec{b}). \end{aligned}$$

Anyway, the upshot is that rotating a vector $\vec{v}$ in a right-handed sense about the axis corresponding to the unit vector $\vec{u}$ through an angle of 2α produces the vector

$$-2(\vec{b} \times \vec{v}) \times \vec{b} + (\vec{b} \times \vec{v})2a + \vec{v},$$

where $a = \cos\alpha$ and $\vec{b} = \vec{u}\sin\alpha$. This is the formula for rotating vectors about arbitrary axes in three-dimensional space.

Recall that the inverse of the quaternion $a + \vec{b}$ is $a - \vec{b}$. In polar form $e^{\vec{u}\alpha}$, inversion is equivalent to negating the angle α; that makes sense, because the inverse of a rotation by 2α should be a rotation by $-2\alpha = 2(-\alpha)$. It is also equivalent to negating $\vec{u}$, which makes sense, because this switches the direction of rotation according to the right-hand rule. So inversion of rotations is easy, when they are expressed as unit quaternions.

Composition of rotations is also easy. The rotation $a + \vec{b}$ followed by the rotation $c + \vec{d}$ is the product $(c + \vec{d})(a + \vec{b})$. We've already seen that in polar form, if the unit vectors $\vec{v}$ and $\vec{u}$ are equal, then multiplication is equivalent to adding angles; that makes sense, since rotating by 2α and then by 2β about the same axis should be equivalent to rotating by $2(\beta + \alpha)$.

Finally, recall that negating a unit quaternion amounts to adding π to its angle:

$$-(\cos\alpha + \vec{u}\sin\alpha) = (-\cos\alpha) + \vec{u}(-\sin\alpha) = \cos(\alpha + \pi) + \vec{u}\sin(\alpha + \pi).$$

The negated quaternion represents a rotation through

$$2(\alpha + \pi) = 2\alpha + 2\pi,$$

which is equivalent to rotating through 2α, since extra multiples of 2π don't matter. So we see that $e^{\vec{u}\alpha}$ and $-e^{\vec{u}\alpha}$ represent the same rotation. It turns out that there are no other unit quaternions that represent this rotation. Hence the group of unit quaternions is "exactly twice as large" as the group of three-dimensional rotations. In particular, 1 and -1 are distinct quaternions, but they both represent the trivial rotation.

In the end, rotation of $\mathbb{R}^3$ by unit quaternions is analogous to rotation of $\mathbb{R}^2$ by unit complex numbers, but not perfectly. The unit complex number $e^{i\alpha}$ acts by multiplication, with the effect of rotating by α, while the unit quaternion $e^{\vec{u}\alpha}$ acts by conjugation, with the effect of rotating by 2α. Unit complex numbers correspond one-to-one to rotations of $\mathbb{R}^2$, while unit quaternions correspond two-to-one to rotations of $\mathbb{R}^3$. These discrepancies are connected to spin and Clifford algebras, but those lie beyond the scope of this paper.

9. Missing Stuff

interpolation: matrices and quaternions
references