

Quaternion Proofs

Kevin

February 6, 2019

Theorem 1. Any unit quaternion $q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ can be written as $q = \cos(\theta) + \mathbf{v} \sin(\theta)$.

Proof. We'll be using the trig formula $\cos^2 \theta + \sin^2 \theta = 1$. Since $\|\mathbf{q}\|^2 = 1$, then $a^2 + b^2 + c^2 + d^2 = 1$. This implies that $a^2 \leq 1$ which means we can write $a = \cos \theta$ and $\sqrt{b^2 + c^2 + d^2} = \sin \theta$ for some $\theta \in \mathbb{R}$. Now since any vector \mathbf{v} can be written as the product of a scalar and a unit-vector, $\mathbf{v} = \|\mathbf{v}\| \frac{\mathbf{v}}{\|\mathbf{v}\|}$, we can express the quaternion vector component as:

$$b\mathbf{i} + c\mathbf{j} + d\mathbf{k} = \sqrt{b^2 + c^2 + d^2} \frac{b\mathbf{i} + c\mathbf{j} + d\mathbf{k}}{\sqrt{b^2 + c^2 + d^2}} = (\sin \theta) \mathbf{v}$$

where \mathbf{v} is a unit vector. Thus, we can write $q = \cos(\theta) + \mathbf{v} \sin(\theta)$. \square

Theorem 2. For a unit-quaternion $q = [s, \mathbf{v}]$ we have that $s^2 + \|\mathbf{v}\|^2 = 1$ and $s^2 - \|\mathbf{v}\|^2 = 2s^2 - 1$.

Proof. The first equation goes without saying. We have that $\|q\|^2 = 1 \implies s^2 + \|\mathbf{v}\|^2 = 1$. For the second equation, we write the vector component in terms of s to obtain: $s^2 - \|\mathbf{v}\|^2 = s^2 - (1 - s^2) = 2s^2 - 1$. \square

Theorem 3. A unit quaternion $q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ can be represented by an orthogonal rotation matrix which can be expressed in different ways. Show that A and B , which both appear on the Wikipedia page, are equivalent:

$$A = \begin{bmatrix} a^2 + b^2 - c^2 - d^2 & 2bc - 2ad & 2bd + 2ac \\ 2bc + 2ad & a^2 - b^2 + c^2 - d^2 & 2cd - 2ab \\ 2bd - 2ac & 2cd + 2ab & a^2 - b^2 - c^2 + d^2 \end{bmatrix}$$
$$B = \begin{bmatrix} 1 - 2(c^2 + d^2) & 2(bc - da) & 2(bd + ca) \\ 2(bc + da) & 1 - 2(b^2 + d^2) & 2(cd - ba) \\ 2(bd - ca) & 2(cd + ba) & 1 - 2(b^2 + c^2) \end{bmatrix}$$

Proof. We'll be relying on the property that the unit quaternion has unit norm, i.e. $a^2 + b^2 + c^2 + d^2 = 1$. All the elements of A and B that are not on the diagonal are obviously equal, so I'll be proving equivalency for the diagonal elements.

First, we can write a and b as a function of c and d: $a^2 + b^2 = 1 - c^2 - d^2$ and then replace in the unit-norm equation to obtain: $a^2 + b^2 - c^2 - d^2 = 1 - c^2 - d^2 - c^2 - d^2 = 1 - 2(c^2 + d^2)$. Second, we can write a and c as $a^2 + c^2 = 1 - b^2 - d^2$ and plug in to obtain $a^2 - b^2 + c^2 - d^2 = 1 - b^2 - d^2 - b^2 - d^2 = 1 - 2(b^2 + d^2)$. The last diagonal element uses a similar manipulation.

From an implementation perspective, the B matrix is more efficient to implement since it involves fewer multiplications. \square

Theorem 4. *In a unit quaternion $q = a + bi + cj + dk$, there must at least one component whose magnitude is at least $\sqrt{1/3} \approx 0.577$.*

Proof. The norm of a unit-quaternion is unitary, i.e. $a^2 + b^2 + c^2 + d^2 = 1$. We can split this analysis into 4 cases: no element is zero, one element is 0, two elements are zero and three elements are zero (we cannot have all elements to zero because it is not a unit quaternion).

If three elements are zero, then it follows that the last component, say c, must respect $c^2 = 1$ which means $c = 1$. Thus its magnitude is at least 0.5.

If two elements are zero, then the worst case scenario is that we split the 1 across both non-zero components equal, say c and d. Then $c^2 = 0.5$ and $d^2 = 0.5$ which means $c = d \approx 0.707$ and again, we have a component whose magnitude is at least 0.5. If we do not split the magnitude 1 equally amongst both, then what will happen is that one of the component's magnitude will decrease but the other one's magnitude will increase above 0.707 which means we still respect the statement.

Finally, if we have 1 element that is zero, then the worst-case scenario is that we split the 1 across three elements in which case we get all three with magnitude $\sqrt{1/3}$. Anything other than this will result in the magnitude of one of the three components decreasing below this number but the magnitude of any of the other two increasing above 0.577 which still evaluates to true. \square

Theorem 5. *To rotate a vector $p = [x, y, z]$ by an angle θ around an arbitrary axis of rotation defined by the unit vector \hat{u} , we construct a quaternion $q = [\cos \frac{1}{2}\theta, \hat{u} \sin \frac{1}{2}\theta]$ and evaluate the quaternion product qpq^{-1} . (Basically, in*

this theorem, I'm trying to show why we use half the angle rather than the full angle for rotation.)

Proof. ... □

Theorem 6. To find the nearest orthogonal matrix to a given matrix M , one uses the singular value decomposition $M = U\Sigma V^T$ to obtain the rotation matrix $R = UV^T$.

Proof. Suppose we have two matrices X and Y and we would like to find a rotation matrix R that closely maps X to Y . Concretely, we can frame this problem as one of minimizing the following energy:

$$\arg \min_R \|RX - Y\|_{\text{Fro}}^2 \text{ s.t. } R^T R = I_{3 \times 3}$$

First, we note the following properties of the trace operator:

$$\begin{aligned} \text{tr}(A^T A) &= \|A\|_{\text{Fro}}^2 \\ \text{tr}(A) &= \text{tr}(A^T) \\ \text{tr}(A + B) &= \text{tr}(A) + \text{tr}(B) \\ \text{tr}(ABC) &= \text{tr}(CAB) = \text{tr}(BCA) \end{aligned}$$

We also note that R is orthogonal which means $R^T R = R R^T = I$. Thus, we can write the following:

$$\begin{aligned} \|RX - Y\|_{\text{Fro}}^2 &= \text{tr}((RX - Y)^T (RX - Y)) \\ &= \text{tr}((X^T R^T - Y^T)(RX - Y)) \\ &= \text{tr}(X^T R^T R X - X^T R^T Y - Y^T R X + Y^T Y) \\ &= \text{tr}(X^T X) - \text{tr}(X^T R^T Y) - \text{tr}(Y^T R X) + \text{tr}(Y^T Y) \\ &= \text{tr}(X^T X) + \text{tr}(Y^T Y) - 2\text{tr}(X^T R^T Y) \end{aligned}$$

We want to minimize the initial energy with respect to R so this means it is equivalent to minimizing $-2\text{tr}(X^T R^T Y)$ which is equivalent to maximizing $\text{tr}(X^T R^T Y) = \text{tr}(Y X^T R^T) = \text{tr}(M R^T)$. As a small aside, you can now see how the original statement with the matrix M is equivalent to this problem, i.e. $M = Y X^T$. Anyway, let's develop the expression some more:

$$\begin{aligned} \text{tr}(X^T R^T Y) &= \text{tr}(Y X^T R^T) \\ &= \text{tr}(U \Sigma V^T R^T) \\ &= \text{tr}(\Sigma V^T R^T U) \\ &= \text{tr}(\Sigma Z) \\ &= \sum_i \sigma_i z_{ii} \end{aligned}$$

We note that $Z = V^T R^T U$ is orthogonal since it is the product of orthogonal matrices. Then the trace of the product of a diagonal matrix and another matrix is just the product of the diagonal entries of the two matrices.

So our goal is to maximize this sum. We note that Z is orthogonal so its columns have unit length. This means that $z_{ii} \leq 1$. The maximal case is when $Z = I$. So we back-solve to obtain:

$$\begin{aligned} Z &= I \\ V^T R^T U &= I \\ R^T &= V U^T \\ R &= (V U^T)^T = U V^T \end{aligned}$$

Thus, we have shown that the nearest orthogonal matrix is

$$\boxed{R = U V^T}$$

□