Quaternions: what are they, and why do we need to know?

Berthold K. P. Horn

Quaternions: what are they, and why do we need to know?

Berthold K. P. Horn*

Computer Science and Artificial Intelligence Laboratory (CSAIL), Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. *Correspondence e-mail: bkph@csail.mit.edu

How should we represent rotations? Orthonormal (3×3) matrices? Euler angles (pitch, yaw and roll)? Axis and angle? Gibbs vector? Pauli (2×2, complex) spin matrices? Cayley–Klein (complex) parameters? Euler–Rodrigues parameters? Hamilton's quaternions? Why are there so many choices (Goldstein, 1980; Varshalovich *et al.*, 1988; Shuster, 1993)? We don't see this variety when it comes to translation: we just use vectors. The big difference is that translations live in a Euclidean vector space. Rotations do not, and so all the nice linear algebra machinery that we use for manipulating translations seems not to apply.

Why does it matter how we represent rotation? Solving problems often depends on finding the 'right' representation. Sometimes a good notation can make it possible to solve a problem in closed form. Without a notion of the 'space of rotations' many problems are hard to formulate and solve.
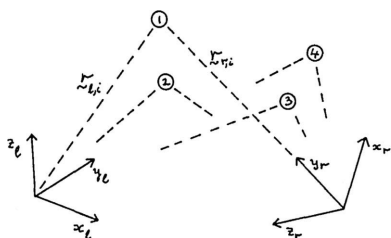
Spoiler alert: unit quaternions provide 'the' way to represent rotations. Why? Unit quaternions allow a clear visualization (see Hanson, 2006) of the space of rotations as the unit sphere S_3 in four dimensions (with antipodal points identified). Unit quaternions make it possible to differentiate an expression 'with respect to rotation'. Unit quaternions make it possible to find extrema of expressions by setting that derivative equal to zero!

Unit quaternions make it easy to compose rotations (unlike, *e.g.*, axis-and-angle notation). Unit quaternions do not suffer from singularities (as do, *e.g.*, Euler angles when two axes line up – see *gimbal lock*). Unit quaternions, while redundant (four parameters for three degrees of freedom), have only one constraint on their components (unlike orthonormal matrices, which have six non-linear constraints on the rows or columns – plus one on the determinant). Unit quaternions make it possible to compute averages, to interpolate, to sample in the space of rotations and to represent densities over orientations. And, as demonstrated in Andrew Hanson's (2020) article in the previous issue of *Acta Crystallographica* Section A, they also make possible insightful illustrations relating to orientations.

Which came first, the vector or the quaternion? Mathematicians extended algebra from real number to complex numbers. We can add, subtract, multiply and divide these. The question that obsessed Hamilton was whether there would be further generalizations to algebras with *three* components. His big breakthrough came when he realized it was *not* possible with three components – but it *was* possible with four (Hamilton, 1844)! That is how the quaternion was born – on 16 October 1843.

Hamilton showed that for some applications in physics it was convenient to split the quaternion into a 'scalar' and a 'vector' part. For a few decades, quaternions were used in physics to represent spatial quantities, often with the scalar part set to zero. Maxwell used both notations in some of his work (Maxwell, 1873). Some got tired of using four components to represent positions in space which clearly only need three, and so lopped off the scalar component and switched to using just vectors. Gibbs and Heaviside were the main propagandizers of this approach (Crowe, 1967).

The price they paid was that there is no algebra for vectors (*e.g.*, no inverses with respect to multiplication). Some critics called vector analysis a 'hermaphrodite monster' because it 'compounded the notations of Hamilton and Grassmann' (Tait, 1867; Prichard, 1998). After an acrimonious war of words, the proponents of vectors finally won, and the rest is history.

Sadly, when quaternions were abandoned, 'the baby was thrown out with the bathwater' so to speak, since the inherent advantages of using unit quaternions to represent rotations were forgotten (Lambek, 1995). In the intervening years, unit quaternions, and their use in solving real problems involving rotations, have been independently rediscovered a number of times in a variety of fields, as shown by Professor Hanson. In crystallography, *e.g.*, the orientation distribution function (ODF) – the volume fraction of grains with certain orientation – can be treated as a function of a unit quaternion (Kunze & Schaeben, 2004; Mason & Schuh, 2008).

The problem of best-fit alignment of points comes up in several disciplines, including aerospace, photogrammetry, machine vision, molecular chemistry, cryo-electron microscopy and crystallography. A simple example of the point alignment problem is illustrated in Fig. 1 (from an old paper: Horn, 1987). Here points on an object are measured by two instruments, each with its own coordinate system. The problem is to determine the transformation between the two systems. An equivalent problem is one where two objects are measured in a common coordinate system. In that case, one is to determine the transformation that brings one into alignment with the other. There are many variants of these problems, *e.g.*, the 'two objects' may be the same object before and after rotation, or two instances (*e.g.*, molecules) from the same object class. In his lead article, Andrew Hanson, for the first time, traces back the parallel development of the elegant quaternion-based solution methods to these problems.

Duplication of effort, without cross reference, is, of course, not unique to the alignment problem, and presents an interesting challenge: How can we detect such parallel efforts across disciplines early enough to avoid a lot of redundant work? What search method can overcome the differences in notation, application, research methodology and underlying philosophy to ferret out such connections?

The repeated rediscovery of important methods is, of course, not uncommon. In the second edition (1973) of *Numerical Methods for Scientists and Engineers*, Hamming starts an introduction to the fast Fourier transform (FFT) with wording that may seem a bit odd:
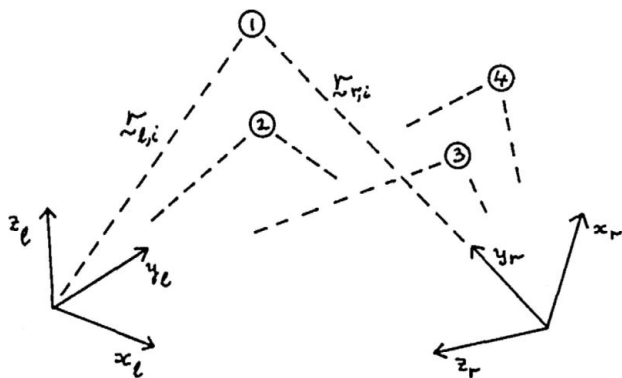
The fast Fourier transform was rediscovered and adequately publicized by Cooley & Tukey (1965), though it had been discovered a number of times before, and was to some extent understood in the general literature.

To his chagrin, apparently nobody had noticed that Hamming had already taught how to do the fast Fourier transform for cyclical data of length 6, 8 and 12 in the first edition of this book (Hamming, 1962). He must have assumed that it would have been obvious how his method could be used for other, larger composite numbers. (The real problem may have been that he did not give his method a snazzy name!) Since then it has become known that FFT-like methods had actually been re-invented numerous times and can be traced back even further, perhaps as far as Gauss (Heideman *et al.*, 1984; Blahut, 1985; Brandt, 2014; Goldstine, 1977; Brendel, 1900*a*). [A relevant aside: Gauss also may have predated Hamilton in discovering quaternions (Altmann, 1989; Brendel, 1900*b*).]

Being unaware of existing solutions of problems can happen to the best of us. In *The Elastica: a Mathematical History*, Raph Levien (2008) writes

In spite of the fairly rich literature available on the elastica, at least one researcher, B. K. P. Horn in 1981, seems to have independently derived the rectangular elastica from the principle of minimizing the strain energy (Horn, 1983), going through an impressive series of derivations, and using the full power of elliptic integral theory, to arrive at exactly the same integral as Bernoulli had derived almost three hundred years previously.

Good thing I didn't end up with a different integral!

Andrew Hanson presents the fascinating story of multiple discoveries of the formulation and solution of the alignment problem. He draws particular attention to those that noticed that closed-form solutions are possible. This is because finding the eigenvalues of a 4×4 matrix involves determining the roots of a quartic, and equations up to fourth order do have closed-form solutions. In the process, he gives a delightful account of the history of the solution of quartics as well. As is common, having a closed-form solution exposes features of the problem not obvious when only numerical solutions are possible. Finally, in his *tour de force*, he further extends the problem from spatial alignment of points to alignment of orientation frames.

Hanson's article should be of great interest to crystallographers, since spatial orientation is an important topic there, and so unit quaternions can play a significant role in, *e.g.*, analysis of crystallographic orientation, crystallographic lattices, crystallographic texture, microstructure and anisotropy (Kunze & Schaeben, 2004; Mason & Schuh, 2008; King, 1996; Ozturk & Rollett, 2016). It is also a most entertaining read.

Figure 1
One version of the absolute orientation problem in photogrammetry. The circled points are measured in two coordinate systems. The task is to find the relationship between the coordinate systems.

References

Altmann, S. L. (1989). *Math. Mag.* **62**, 306.
Blahut, R. E. (1985). *Fast Algorithms for Digital Signal Processing*. Springer.

scientific commentaries

Brandt, A. (2014). *Special Topics in Structural Dynamics*, edited by G. Foss & C. Niezrecki, Vol. 6, pp. 33–36. Conference Proceedings of the Society for Experimental Mechanics Series. Springer.

Brendel, M. (1900a). Editor. *Carl Friedrich Gauss Werke*, Vol. 3, pp. 265–303. Göttingen: Königlichen Gesellschaft der Wissenschaften.

Brendel, M. (1900b). Editor. *Carl Friedrich Gauss Werke*, Vol. 8, pp. 357–361. Göttingen: Königlichen Gesellschaft der Wissenschaften.

Cooley, J. W. & Tukey, J. W. (1965). *Math. Comp.* **19**, 297–301.

Crowe, M. J. (1967). *A History of Vector Analysis: the Evolution of the Idea of a Vectorial System*. Dover.

Goldstein, H. (1980). *Classical Mechanics*, 2nd ed., Section 4-5, pp. 148–158. Reading, MA: Addison-Wesley.

Goldstine, H. H. (1977). *A History of Numerical Analysis from the 16th through the 19th Century*, pp. 249–253. New York: Springer.

Hamilton, W. R. (1844). *Philos. Mag.* **25**, 489–495.

Hamming, R. W. (1962). *Numerical Methods for Scientists and Engineers*. New York: McGraw-Hill.

Hamming, R. W. (1973). *Numerical Methods for Scientists and Engineers*, 2nd ed. New York: McGraw-Hill.

Hanson, A. J. (2006). *Visualizing Quaternions*. San Francisco: Morgan-Kaufmann/Elsevier.

Hanson, A. J. (2020). *Acta Cryst.* A**76**, 432–457.

Heideman, M. T., Johnson, D. H. & Burrus, C. S. (1984). *IEEE Signal Process. Mag.* **1**, 14–21.

Horn, B. K. P. (1983). *ACM Trans. Math. Softw.* **9**, 441–460.

Horn, B. K. P. (1987). *J. Opt. Soc. Am. A*, **4**, 629–642.

King, R. B. (1996). *Croat. Chem. Acta*, **69**, 805–812.

Kunze, K. & Schaeben, H. (2004). *Math. Geol.* **36**, 917–943.

Lambek, J. (1995). *Math. Intelligencer*, **17**, 7–15.

Levien, R. (2008). *The Elastica: a Mathematical History*. Report UCB/EECS-2008-103, EECS Department, University of California, Berkeley, USA. http://levien.com/phd/elastica_hist.pdf.

Mason, J. K. & Schuh, C. A. (2008). *Acta Mater.* **56**, 6141–6155.

Maxwell, J. C. (1873). *A Treatise on Electricity and Magnetism*. Oxford: Clarendon Press.

Ozturk, T. & Rollett, A. (2016). http://pajarito.materials.cmu.edu/rollett/27750/L20-Descriptions-of-Orientation-TugceOzturk_27750-12Apr16.pptx.

Prichard, C. (1998). *Math. Gaz.* **82**, No. 494.

Shuster, M. D. (1993). *J. Astronaut. Sci.* **41**, 439–517.

Tait, P. G. (1867). *An Elementary Treatise on Quaternions*. Oxford: Clarendon Press.

Varshalovich, D. A., Moskalev, A. N. & Khersonskii, V. K. (1988). *Quantum Theory of Angular Momentum*, Section 1.4.3, pp. 24–27. Singapore: World Scientific.

electronic reprint