



The quaternion-based spatial-coordinate and orientation-frame alignment problems

Andrew J. Hanson*

Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, Indiana, USA.

*Correspondence e-mail: hansonaj@indiana.edu

Received 4 September 2018

Accepted 25 February 2020

Edited by S. J. L. Billinge, Columbia University, USA

Keywords: data alignment; spatial-coordinate alignment; orientation-frame alignment; quaternions; quaternion frames; quaternion eigenvalue methods.

Andrew J. Hanson is an Emeritus Professor of Computer Science at Indiana University. He earned a bachelor's degree in Chemistry and Physics from Harvard University in 1966 and a PhD in Theoretical Physics from MIT in 1971. His interests range from general relativity to computer graphics, artificial intelligence and bioinformatics; he is particularly concerned with applications of quaternions and with exploitation of higher-dimensional graphics for the visualization of complex scientific contexts such as Calabi–Yau spaces. He is the co-discoverer of the Eguchi–Hanson ‘gravitational instanton’ Einstein metric (1978), author of *Visualizing Quaternions* (Elsevier, 2006), and designer of the iPhone apps *4Dice* and *4DRoom* (2012) for interacting with four-dimensional virtual reality.

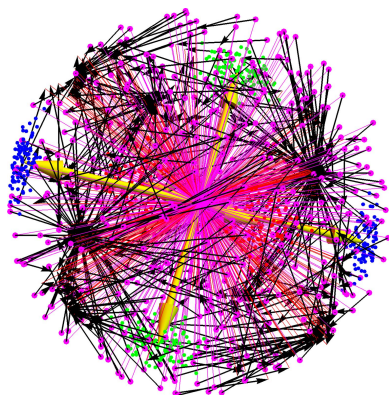
Supporting information: this article has supporting information at journals.iucr.org/a

The general problem of finding a global rotation that transforms a given set of spatial coordinates and/or orientation frames (the ‘test’ data) into the best possible alignment with a corresponding set (the ‘reference’ data) is reviewed. For 3D point data, this ‘orthogonal Procrustes problem’ is often phrased in terms of minimizing a root-mean-square deviation (RMSD) corresponding to a Euclidean distance measure relating the two sets of matched coordinates. This article focuses on quaternion eigensystem methods that have been exploited to solve this problem for at least five decades in several different bodies of scientific literature, where they were discovered independently. While numerical methods for the eigenvalue solutions dominate much of this literature, it has long been realized that the quaternion-based RMSD optimization problem can also be solved using exact algebraic expressions based on the form of the quartic equation solution published by Cardano in 1545; focusing on these exact solutions exposes the structure of the entire eigensystem for the traditional 3D spatial-alignment problem. The structure of the less-studied orientation-data context is then explored, investigating how quaternion methods can be extended to solve the corresponding 3D quaternion orientation-frame alignment (QFA) problem, noting the interesting equivalence of this problem to the rotation-averaging problem, which also has been the subject of independent literature threads. The article concludes with a brief discussion of the combined 3D translation–orientation data alignment problem. Appendices are devoted to a tutorial on quaternion frames, a related quaternion technique for extracting quaternions from rotation matrices and a review of quaternion rotation-averaging methods relevant to the orientation-frame alignment problem. The supporting information covers novel extensions of quaternion methods to the 4D Euclidean spatial-coordinate alignment and 4D orientation-frame alignment problems, some miscellaneous topics, and additional details of the quartic algebraic eigenvalue problem.

1. Context

Aligning matched sets of spatial point data is a universal problem that occurs in a wide variety of applications. In addition, generic objects such as protein residues, parts of composite object models, satellites, cameras, or camera-calibrating reference objects are not only located at points in three-dimensional space, but may also need 3D orientation frames to describe them effectively for certain applications. We are therefore led to consider both the Euclidean spatial-coordinate alignment problem and the orientation-frame alignment problem on the same footing.

Our purpose in this article is to review, and in some cases to refine, clarify and extend, the possible quaternion-based approaches to the optimal alignment problem for matched sets of translated and/or rotated objects in 3D space, which could be referred to in its most generic sense as the ‘generalized orthogonal Procrustes problem’ (Golub & van Loan, 1983). We also devote some attention to identifying the surprising



breadth of domains and literature where the various approaches, including particularly quaternion-based methods, have appeared; in fact the number of times in which quaternion-related methods have been described independently without cross-disciplinary references is rather interesting, and exposes some challenging issues that scientists, including the author, have faced in coping with the wide dispersion of both historical and modern scientific literature relevant to these subjects.

We present our study on two levels. The first level, the present main article, is devoted to a description of the 3D spatial and orientation alignment problems, emphasizing quaternion methods, with an historical perspective and a moderate level of technical detail that strives to be accessible. The second level, comprising the supporting information, treats novel extensions of the quaternion method to the 4D spatial and orientation alignment problems, along with many other technical topics, including analysis of algebraic quartic eigenvalue solutions and numerical studies of the applicability of certain common approximations and methods.

In the following, we first review the diverse bodies of literature regarding the extraction of 3D rotations that optimally align matched pairs of Euclidean point data sets. It is important for us to remark that we have repeatedly become aware of additional literature in the course of this work and it is entirely possible that other worthy references have been overlooked; if so, we apologize for any oversights and hope that the literature that we have found to review will provide an adequate context for the interested reader. We then introduce our own preferred version of the quaternion approach to the spatial-alignment problem, often described as the root-mean-square-deviation (RMSD) minimization problem, and we will adopt that terminology when convenient; our intent is to consolidate a range of distinct variants in the literature into one uniform treatment, and, given the wide variations in symbolic notation and terminology, here we will adopt terms and conventions that work well for us personally. Following a technical introduction to quaternions, we treat the quaternion-based 3D spatial-alignment problem itself. Next we introduce the quaternion approach to the 3D orientation-frame alignment (QFA) problem in a way that parallels the 3D spatial problem, and note its equivalence to quaternion-frame averaging methods. We conclude with a brief analysis of the 6-degree-of-freedom (6DOF) problem, combining the 3D spatial and 3D orientation-frame measures. Appendices include treatments of the basics of quaternion orientation frames, an elegant method that extracts a quaternion from a numerical 3D rotation matrix and the generalization of that method to compute averages of rotations.

2. Summary of spatial-alignment problems, known solutions and historical contexts

2.1. The problem, standard solutions and the quaternion method

The fundamental problem we will be concerned with arises when we are given a well behaved $D \times D$ matrix E and we

wish to find the optimal D -dimensional proper orthogonal matrix R_{opt} that maximizes the measure $\text{tr}(R \cdot E)$. This is equivalent to the RMSD problem, which seeks a global rotation R that rotates an ordered set of point test data X in such a way as to minimize the squared Euclidean differences relative to a matched reference set Y . We will find below that E corresponds to the cross-covariance matrix of the pair (X, Y) of N columns of D -dimensional vectors, namely $E = X \cdot Y^T$, though we will look at cases where E could have almost any origin.

One solution to this problem in any dimension D uses the decomposition of the general matrix E into an orthogonal matrix U and a symmetric matrix S that takes the form $E = U \cdot S = U \cdot (E^T \cdot E)^{1/2}$, giving $R_{\text{opt}} = U^{-1} = (E^T \cdot E)^{1/2} \cdot E^{-1}$; note that there exist several equivalent forms [see, e.g., Green (1952) and Horn *et al.* (1988)]. General solutions may also be found using singular-value-decomposition (SVD) methods, starting with the decomposition $E = U \cdot S \cdot V^T$, where S is now diagonal and U and V are orthogonal matrices, to give the result $R_{\text{opt}} = V \cdot D \cdot U^T$, where D is the identity matrix up to a possible sign in one element [see, e.g., Kabsch (1976, 1978), Golub & van Loan (1983) and Markley (1988)].

In addition to these general methods based on traditional linear algebra approaches, a significant literature exists for three dimensions that exploits the relationship between 3D rotation matrices and quaternions, and rephrases the task of finding R_{opt} as a *quaternion eigensystem* problem. This approach notes that, using the quadratic quaternion form $R(q)$ for the rotation matrix, one can rewrite $\text{tr}(R \cdot E) \rightarrow q \cdot M(E) \cdot q$, where the *profile matrix* $M(E)$ is a traceless, symmetric 4×4 matrix consisting of linear combinations of the elements of the 3×3 matrix E . Finding the largest eigenvalue ϵ_{opt} of $M(E)$ determines the optimal quaternion eigenvector q_{opt} and thus the solution $R_{\text{opt}} = R(q_{\text{opt}})$. The quaternion framework will be our main topic here.

2.2. Historical literature overview

Although our focus is the quaternion eigensystem context, we first note that one of the original approaches to the RMSD task exploited the singular-value decomposition directly to obtain an optimal rotation matrix. This solution appears to date at least from 1966 in Schönemann's thesis (Schönemann, 1966) and possibly Cliff (1966) later in the same journal issue; Schönemann's work is chosen for citation, for example, in the earliest editions of Golub & van Loan (1983). Applications of the SVD to alignment in the aerospace literature appear, for example, in the context of Wahba's problem (Wikipedia, 2018b; Wahba, 1965) and are used explicitly, e.g., in Markley (1988), while the introduction of the SVD for the alignment problem in molecular chemistry is generally attributed to Kabsch (Wikipedia, 2018a; Kabsch, 1976), and in machine vision Arun *et al.* (1987) is often cited.

We believe that the quaternion eigenvalue approach itself was first noticed around 1968 by Davenport (Davenport, 1968) in the context of Wahba's problem, rediscovered in 1983 by Hebert and Faugeras (Hebert, 1983; Faugeras & Hebert, 1983,

1986) in the context of machine vision, and then found independently a third time in 1986 by Horn (Horn, 1987).

An alternative quaternion-free approach by Horn *et al.* (1988) with the optimal rotation of the form $R_{opt} = (E^T \cdot E)^{1/2} \cdot E^{-1}$ appeared in 1988, but this basic form was apparently known elsewhere as early as 1952 (Green, 1952; Gibson, 1960).

Much of the recent activity has occurred in the context of the molecular alignment problem, starting from a basic framework put forth by Kabsch (1976, 1978). So far as we can determine, the matrix eigenvalue approach to molecular alignment was introduced in 1988 without actually mentioning quaternions by name in Diamond (1988) and refined to specifically incorporate quaternion methods in 1989 by Kearsley (1989). In 1991 Kneller (Kneller, 1991) independently described a version of the quaternion-eigenvalue-based approach that is widely cited as well. A concise and useful review can be found in Flower (1999), in which the contributions of Schönemann, Faugeras and Hebert, Horn, Diamond, and Kearsley are acknowledged and all cited in the same place. A graphical summary of the discovery chronology in various domains is given in Fig. 1. Most of these treatments mention using numerical methods to find the optimal eigenvalue, though several references, starting with Horn (1987), point out that 16th-century algebraic methods for solving the quartic polynomial characteristic equation, discussed in the next section, could also be used to determine the eigenvalues.

In our treatment we will study the explicit form of these algebraic solutions for the 3D problem (and also for 4D in the supporting information), taking advantage of several threads of the literature.

2.3. Historical notes on the quartic

The actual solution to the quartic equation, and thus the solution of the characteristic polynomial of the 4D eigen-system of interest to us, was first published in 1545 by Gerolamo Cardano (Wikipedia, 2019) in his book *Ars Magna*. The intellectual history of this fact is controversial and narrated with varying emphasis in diverse sources. It seems generally agreed upon that Cardano's student Lodovico Ferrari was the first to discover the basic method for solving the quartic in 1540, but his technique was incomplete as it only reduced the problem to the cubic equation, for which no solution was publicly known at that time, and that apparently prevented him from publishing it. The complication appears to be that Cardano had actually learned of a method for solving the cubic already in 1539 from Niccolò Fontana Tartaglia (legendarily in the form of a poem), but had been sworn to secrecy, and so could not reveal the final explicit step needed to complete Ferrari's implicit solution. Where it gets controversial is that at some point between 1539 and 1545, Cardano learned that Scipione del Ferro had found the same cubic solution as the one of Tartaglia that he had sworn not to

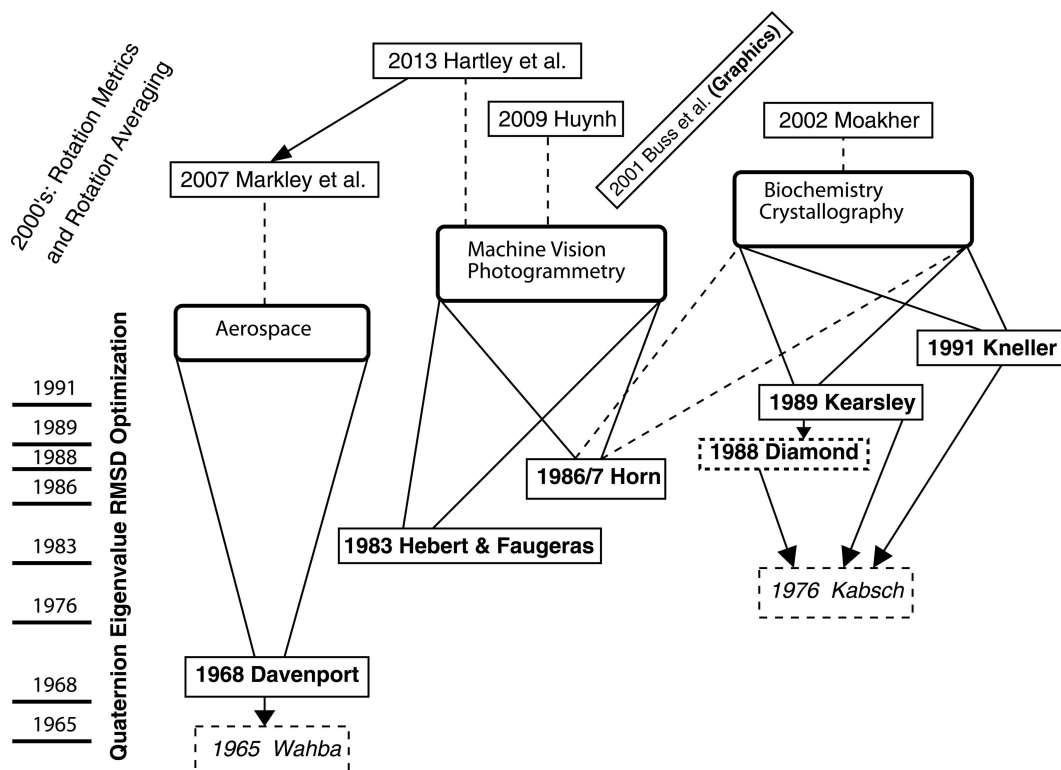


Figure 1

The quaternion eigensystem method for computing the optimal rotation matching two spatial data sets was discovered independently and published without cross-references in at least three distinct literatures. Downward arrows point to the introduction of the abstract problem and upward rays indicate domains of publications specifically citing the quaternion method. Horn eventually appeared routinely in the crystallography citations, and reviews such as that by Flower (1999) introduced multiple cross-field citations. Several fields have included activity on quaternion-related rotation metrics and rotation averaging with varying degrees of cross-field awareness.

reveal, and furthermore that del Ferro had discovered his solution before Tartaglia did. Cardano interpreted that fact as releasing him from his oath of secrecy (which Tartaglia did not appreciate), allowing him to publish the complete solution to the quartic, incorporating the cubic solution into Ferrari's result. Sources claiming that Cardano 'stole' Ferrari's solution may perhaps be exaggerated, since Ferrari did not have access to the cubic equations and Cardano did not conceal his sources; exactly who 'solved' the quartic is thus philosophically complicated, but Cardano does seem to be the one who combined the multiple threads needed to express the equations as a single complete formula.

Other interesting observations were made later, for example by Descartes in 1637 (Descartes, 1637) and in 1733 by Euler (Euler, 1733; Bell, 2008). For further descriptions, one may consult, *e.g.*, Abramowitz & Stegun (1970) and Boyer & Merzbach (1991), as well as the narratives in Weisstein (2019*a,b*). Additional amusing pedagogical investigations of the historical solutions may be found in several expositions by Nickalls (1993, 2009).

2.4. Further literature

A very informative treatment of the features of the quaternion eigenvalue solutions was given by Coutsias, Seok and Dill in 2004, and expanded in 2019 (Coutsias *et al.*, 2004; Coutsias & Wester, 2019). Coutsias *et al.* not only take on a thorough review of the quaternion RMSD method, but also derive the complete relationship between the linear algebra of the SVD method and the quaternion eigenvalue system; furthermore, they exhaustively enumerate the special cases involving mirror geometries and degenerate eigenvalues that may appear rarely, but must be dealt with on occasion. Efficiency is also an area of potential interest, and Theobald *et al.* in Theobald (2005) and in Liu *et al.* (2010) argue that among the many variants of numerical methods that have been used to compute the optimal quaternion eigenvalues, Horn's original proposal to use Newton's method directly on the characteristic equations of the relevant eigenvalue systems may well be the best approach.

There is also a rich literature dealing with distance measures among representations of rotation frames themselves, some dealing directly with the properties of distances computed with rotation matrices or quaternions, *e.g.* Huynh (2009), and others combining discussion of the distance measures with associated applications such as rotation averaging or finding 'rotational centers of mass', *e.g.* Brown & Worsey (1992), Park & Ravani (1997), Buss & Fillmore (2001), Moakher (2002), Markley *et al.* (2007), and Hartley *et al.* (2013). The specific computations explored below in Section 7 on optimal alignment of matched pairs of orientation frames make extensive use of the quaternion-based and rotation-based measures discussed in these treatments. In the appendices, we review the details of some of these orientation-frame-based applications.

3. Introduction

We explore the problem of finding global rotations that optimally align pairs of corresponding lists of spatial and/or orientation data. This issue is significant in diverse application domains. Among these are aligning spacecraft (see, *e.g.*, Wahba, 1965; Davenport, 1968; Markley, 1988; and Markley & Mortari, 2000), obtaining correspondence of registration points in 3D model matching (see, *e.g.*, Faugeras & Hebert, 1983, 1986), matching structures in aerial imagery (see, *e.g.*, Horn, 1987; Horn *et al.*, 1988; Huang *et al.*, 1986; Arun *et al.*, 1987; Umeyama, 1991; and Zhang, 2000), and alignment of matched molecular and biochemical structures (see, *e.g.*, Kabsch, 1976, 1978; McLachlan, 1982; Lesk, 1986; Diamond, 1988; Kearsley, 1989, 1990; Kneller, 1991; Coutsias *et al.*, 2004; Theobald, 2005; Liu *et al.*, 2010; and Coutsias & Wester, 2019). A closely related task is the alignment of multiple sets of 3D range data, for example in digital-heritage applications (Levoy *et al.*, 2000); the widely used iterative closest point (ICP) algorithm [see, *e.g.*, Chen & Medioni (1991), Besl & McKay (1992) and Bergevin *et al.* (1996), as well as Rusinkiewicz & Levoy (2001) and Nüchter *et al.* (2007)] explicitly incorporates standard alignment methods in individual steps with known correspondences.

We note in particular the several alternative approaches of Schönemann (1966), Faugeras & Hebert (1983), Horn (1987) and Horn *et al.* (1988) that in principle produce the same optimal global rotation to solve a given alignment problem. While the SVD and $(E^T \cdot E)^{1/2} \cdot E^{-1}$ methods apply to any dimension, here we will critically examine the quaternion eigensystem decomposition approach that applies only to the 3D and 4D spatial-coordinate alignment problems, along with the extensions to the 3D and 4D orientation-frame alignment problems. Starting from the quartic algebraic equations for the quaternion eigensystem arising in our optimization problem, we direct attention to the elegant exact algebraic forms of the eigenvalue solutions appropriate for these applications. (For brevity, the more complicated 4D treatment is deferred to the supporting information.)

Our extension of the quaternion approach to *orientation data* exploits the fact that collections of 3D orientation frames can *themselves* be expressed as quaternions, *e.g.* amino-acid 3D orientation frames written as quaternions (see Hanson & Thakur, 2012), and we will refer to the corresponding 'quaternion-frame alignment' task as the QFA problem. Various proximity measures for such orientation data have been explored in the literature (see, *e.g.*, Park & Ravani, 1997; Moakher, 2002; Huynh, 2009; and Huggins, 2014*a*), and the general consensus is that the most rigorous measure minimizes the sums of squares of geodesic arc lengths between pairs of quaternions. This ideal QFA proximity measure is highly nonlinear compared to the analogous spatial RMSD measure, but fortunately there is an often-justifiable linearization, the chord angular distance measure. We present several alternative solutions exploiting this approximation that closely parallel our spatial RMSD formulation, and point out the relationship to the rotation-averaging problem.

In addition, we analyze the problem of optimally aligning combined 3D spatial and quaternion 3D-frame-triad data, a 6DOF task that is relevant to studying molecules with composite structure as well as to some gaming and robotics contexts. Such rotational–translational measures have appeared in both the computer vision and the molecular entropy literature [see, *e.g.*, the dual-quaternion approach of Walker *et al.* (1991) as well as Huggins (2014*b*) and Fogolari *et al.* (2016)]; after some confusion, it has been recognized that the spatial and rotational measures are dimensionally incompatible, and either they must be optimized independently, or an arbitrary context-dependent scaling parameter with the dimension of length must appear in any combined measure for the RMSD+QFA problem.

In the following, we organize our thoughts by first summarizing the fundamentals of quaternions, which will be our main computational tool. We next introduce the measures that underlie the general spatial alignment problem, then restrict our attention to the quaternion approach to the 3D problem, emphasizing a class of exact algebraic solutions that can be used as an alternative to the traditional numerical methods. Our quaternion approach to the 3D orientation-frame triad alignment problem is presented next, along with a discussion of the combined spatial–rotational problem. Appendices provide an alternative formulation of the 3D RMSD optimization equations, a tutorial on the quaternion orientation-frame methodology, and a summary of the method of Bar-Itzhack (2000) for obtaining the corresponding quaternion from a numerical 3D rotation matrix, along with a treatment of the closely related quaternion-based rotation averaging problem.

In the supporting information, we extend all of our 3D results to 4D space, exploiting quaternion pairs to formulate the 4D spatial-coordinate RMSD alignment and 4D orientation-based QFA methods. We expose the relationship between these quaternion methods and the singular-value decomposition, and extend the 3D Bar-Itzhack approach to 4D, showing how to find the pair of quaternions corresponding to any numerical 4D rotation matrix. Other sections of the supporting information explore properties of the RMSD problem for 2D data and evaluate the accuracy of our 3D orientation-frame alignment approximations, as well as studying and evaluating the properties of combined measures for aligning spatial-coordinate and orientation-frame data in 3D. An appendix is devoted to further details of the quartic equations and forms of the algebraic solutions related to our eigenvalue problems.

4. Foundations of quaternions

For the purposes of this paper, we take a quaternion to be a point $q = (q_0, q_1, q_2, q_3) = (q_0, \mathbf{q})$ in 4D Euclidean space with unit norm, $q \cdot q = 1$, and so geometrically it is a point on the unit 3-sphere \mathbf{S}^3 [see, *e.g.*, Hanson (2006) for further details about quaternions]. The first term, q_0 , plays the role of a real number and the last three terms, denoted as a 3D vector \mathbf{q} , play the role of a generalized imaginary number, and so are

treated differently from the first: in particular the conjugation operation is taken to be $\bar{q} = (q_0, -\mathbf{q})$. Quaternions possess a multiplication operation denoted by \star and defined as follows:

$$q \star p = Q(q) \cdot p = \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{bmatrix} \cdot \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix} = (q_0 p_0 - \mathbf{q} \cdot \mathbf{p}, q_0 \mathbf{p} + p_0 \mathbf{q} + \mathbf{q} \times \mathbf{p}), \quad (1)$$

where the orthonormal matrix $Q(q)$ expresses a form of quaternion multiplication that can be useful. Note that the orthonormality of $Q(q)$ means that quaternion multiplication of p by q literally produces a rotation of p in 4D Euclidean space.

Choosing exactly one of the three imaginary components in both q and p to be nonzero gives back the classic complex algebra $(q_0 + iq_1)(p_0 + ip_1) = (q_0 p_0 - q_1 p_1) + i(q_0 p_1 + p_0 q_1)$, so there are three copies of the complex numbers embedded in the quaternion algebra; the difference is that in general the final term $\mathbf{q} \times \mathbf{p}$ changes sign if one reverses the order, making the quaternion product order-dependent, unlike the complex product. Nevertheless, like complex numbers, the quaternion algebra satisfies the non-trivial ‘multiplicative norm’ relation

$$\|q\| \|p\| = \|q \star p\|, \quad (2)$$

where $\|q\|^2 = q \cdot q = \Re(q \star \bar{q})$, *i.e.* quaternions are one of the four possible Hurwitz algebras (real, complex, quaternion and octonion).

Quaternion triple products obey generalizations of the 3D vector identities $A \cdot (B \times C) = B \cdot (C \times A) = C \cdot (A \times B)$, along with $A \times B = -B \times A$. The corresponding quaternion identities, which we will need in Section 7, are

$$r \cdot (q \star p) = q \cdot (r \star \bar{p}) = \bar{p} \cdot (\bar{r} \star q) = \bar{r} \cdot (\bar{p} \star \bar{q}), \quad (3)$$

where the complex-conjugate entries are the natural consequences of the sign changes occurring only in the 3D part.

It can be shown that quadratically conjugating a vector $\mathbf{x} = (x, y, z)$, written as a purely ‘imaginary’ quaternion $(0, \mathbf{x})$ (with only a 3D part), by quaternion multiplication is isomorphic to the construction of a 3D Euclidean rotation $R(q)$ generating all possible elements of the special orthogonal group $\mathbf{SO}(3)$. If we compute

$$q \star (c, x, y, z) \star \bar{q} = (c, R(q) \cdot \mathbf{x}), \quad (4)$$

we see that only the purely imaginary part is affected, whether or not the arbitrary real constant $c = 0$. The result of collecting coefficients of the vector term is a proper orthonormal 3D rotation matrix quadratic in the quaternion elements that takes the form

$$\left. \begin{aligned} R_{ij}(q) &= \delta_{ij}(q_0^2 - \mathbf{q}^2) + 2q_i q_j - 2\epsilon_{ijk} q_0 q_k \\ R(q) &= \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2q_1 q_2 - 2q_0 q_3 & 2q_1 q_3 + 2q_0 q_2 \\ 2q_1 q_2 + 2q_0 q_3 & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2q_2 q_3 - 2q_0 q_1 \\ 2q_1 q_3 - 2q_0 q_2 & 2q_2 q_3 + 2q_0 q_1 & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \end{aligned} \right\} \quad (5)$$

with determinant $\det R(q) = (q \cdot q)^3 = +1$. The formula for $R(q)$ is technically a two-to-one mapping from quaternion space to the 3D rotation group because $R(q) = R(-q)$; changing the sign of the quaternion preserves the rotation matrix. Note also that the identity quaternion $q_{\text{ID}} = (1, 0, 0, 0) \equiv q \star \bar{q}$ corresponds to the identity rotation matrix, as does $-q_{\text{ID}} = (-1, 0, 0, 0)$. The 3×3 matrix $R(q)$ is fundamental not only to the quaternion formulation of the spatial RMSD alignment problem, but will also be essential to the QFA orientation-frame problem because the *columns* of $R(q)$ are exactly the needed quaternion representation of the *frame triad* describing the orientation of a body in 3D space, *i.e.* the columns are the vectors of the frame's local x , y and z axes relative to an initial identity frame.

Multiplying a quaternion p by the quaternion q to get a new quaternion $p' = q \star p$ simply *rotates* the 3D frame corresponding to p by the matrix equation (5) written in terms of q . This has non-trivial implications for 3D rotation matrices, for which quaternion multiplication corresponds exactly to multiplication of two *independent* 3×3 orthogonal rotation matrices, and we find that

$$R(q \star p) = R(q) \cdot R(p). \quad (6)$$

This collapse of repeated rotation matrices to a single rotation matrix with multiplied quaternion arguments can be continued indefinitely.

If we choose the following specific 3-variable parameterization of the quaternion q preserving $q \cdot q = 1$,

$$q = (\cos(\theta/2), \hat{n}_1 \sin(\theta/2), \hat{n}_2 \sin(\theta/2), \hat{n}_3 \sin(\theta/2)) \quad (7)$$

(with $\hat{n} \cdot \hat{n} = 1$), then $R(q) = R(\theta, \hat{n})$ is precisely the 'axis-angle' 3D spatial rotation by an angle θ leaving the direction \hat{n} fixed, so \hat{n} is the lone real eigenvector of $R(q)$.

4.1. The slerp

Relationships among quaternions can be studied using the *slerp*, or 'spherical linear interpolation' (Shoemake, 1985; Jupp & Kent, 1987), which smoothly parameterizes the points on the shortest geodesic quaternion path between two constant (unit) quaternions, q_0 and q_1 , as

$$\text{slerp}(q_0, q_1, s) \equiv q(s)[q_0, q_1] = q_0 \frac{\sin((1-s)\phi)}{\sin \phi} + q_1 \frac{\sin(s\phi)}{\sin \phi}. \quad (8)$$

Here $\cos \phi = q_0 \cdot q_1$ defines the angle ϕ between the two given quaternions, while $q(s=0) = q_0$ and $q(s=1) = q_1$. The 'long' geodesic can be obtained for $1 \leq s \leq 2\pi/\phi$. For small ϕ , this reduces to the standard linear interpolation $(1-s)q_0 + sq_1$. The unit norm is preserved, $q(s) \cdot q(s) = 1$ for all s , so $q(s)$ is always a valid quaternion and $R(q(s))$ defined by equation (5) is always a valid 3D rotation matrix. We note that one can formally write equation (8) as an exponential of the form $q_0 \star (\bar{q}_0 \star q_1)^s$, but since this requires computing a logarithm and an exponential whose most efficient reduction to a practical computer program is equation (8), this is mostly of pedagogical interest.

In the following we will make little further use of the quaternion's algebraic properties, but we will extensively exploit equation (5) to formulate elegant approaches to RMSD problems, along with employing equation (8) to study the behavior of our data under smooth variations of rotation matrices.

4.2. Remark on 4D

Our fundamental formula equation (5) can be extended to four Euclidean dimensions by choosing two *distinct* quaternions in equation (4), producing a 4D Euclidean rotation matrix. Analogously to 3D, the columns of this matrix correspond to the axes of a 4D Euclidean orientation frame. The non-trivial details of the quaternion approach to aligning both 4D spatial-coordinate and 4D orientation-frame data are given in the supporting information.

5. Reviewing the 3D spatial-alignment RMSD problem

We now review the basic ideas of spatial data alignment, and then specialize to 3D (see, *e.g.*, Wahba, 1965; Davenport, 1968; Markley, 1988; Markley & Mortari, 2000; Kabsch, 1976, 1978; McLachlan, 1982; Lesk, 1986; Faugeras & Hebert, 1983; Horn, 1987; Huang *et al.*, 1986; Arun *et al.*, 1987; Diamond, 1988; Kearsley, 1989, 1990; Umeyama, 1991; Kneller, 1991; Coutsias *et al.*, 2004; and Theobald, 2005). We will then employ quaternion methods to reduce the 3D spatial-alignment problem to the task of finding the optimal quaternion eigenvalue of a certain 4×4 matrix. This is the approach we have discussed in the introduction, and it can be solved using numerical or algebraic eigensystem methods. In Section 6 below, we will explore in particular the classical quartic equation solutions for the exact algebraic form of the entire four-part eigensystem, whose optimal eigenvalue and its quaternion eigenvector produce the optimal global rotation solving the 3D spatial-alignment problem.

5.1. Aligning matched data sets in Euclidean space

We begin with the general least-squares form of the RMSD problem, which is solved by minimizing the optimization measure over the space of rotations, which we will convert to an optimization over the space of unit quaternions. We take as input one data array with N columns of D -dimensional points $\{y_k\}$ as the *reference* structure and a second array of N columns of *matched* points $\{x_k\}$ as the *test* structure. Our task is to rotate the latter in space by a global $\mathbf{SO}(D)$ rotation matrix R_D to achieve the minimum value of the cumulative quadratic distance measure

$$\mathbf{S}_D = \sum_{k=1}^N \|R_D \cdot x_k - y_k\|^2. \quad (9)$$

We assume, as is customary, that any overall translational components have been eliminated by displacing both data sets to their centers of mass (see, *e.g.*, Faugeras & Hebert, 1983; Coutsias *et al.*, 2004). When this measure is minimized with

respect to the rotation R_D , the optimal R_D will rotate the test set $\{x_k\}$ to be as close as possible to the reference set $\{y_k\}$. Here we will focus on 3D data sets (and, in the supporting information, 4D data sets) because those are the dimensions that are easily adaptable to our targeted quaternion approach. In 3D, our least-squares measure equation (9) can be converted directly into a quaternion optimization problem using the method of Hebert and Faugeras detailed in Appendix A.

Remark: Clifford algebras may support alternative methods as well as other approaches to higher dimensions [see, e.g., Havel & Najfeld (1994) and Buchholz & Sommer (2005)].

5.2. Converting from least-squares minimization to cross-term maximization

We choose from here onward to focus on an equivalent method based on expanding the measure given in equation (9), removing the constant terms, and recasting the RMSD least-squares minimization problem as the task of maximizing the surviving cross-term expression. This takes the general form

$$\Delta_D = \sum_{k=1}^N (R_D \cdot x_k) \cdot y_k = \sum_{a=1, b=1}^D [R_D]_{ba} E_{ab} = \text{tr } R_D \cdot E, \quad (10)$$

where

$$E_{ab} = \sum_{k=1}^N [x_k]_a [y_k]_b = [\mathbf{X} \cdot \mathbf{Y}^T]_{ab} \quad (11)$$

is the *cross-covariance matrix* of the data, $[x_k]$ denotes the k th column of \mathbf{X} and the range of the indices (a, b) is the spatial dimension D .

5.3. Quaternion transformation of the 3D cross-term form

We now restrict our attention to the 3D *cross-term form* of equation (10) with pairs of 3D point data related by a proper rotation. The key step is to substitute equation (5) for $R(q)$ into equation (10) and pull out the terms corresponding to pairs of components of the quaternions q . In this way the 3D expression is transformed into the 4×4 matrix $M(E)$ sandwiched between two identical quaternions (not a conjugate pair) of the form

$$\Delta(q) = \text{tr } R(q) \cdot E = (q_0, q_1, q_2, q_3) \cdot M(E) \cdot (q_0, q_1, q_2, q_3)^T \equiv q \cdot M(E) \cdot q. \quad (12)$$

Here $M(E)$ is the traceless, symmetric 4×4 matrix

$$M(E) = \begin{bmatrix} E_{xx} + E_{yy} + E_{zz} & E_{yz} - E_{zy} & E_{zx} - E_{xz} & E_{xy} - E_{yx} \\ E_{yz} - E_{zy} & E_{xx} - E_{yy} - E_{zz} & E_{xy} + E_{yx} & E_{zx} + E_{xz} \\ E_{zx} - E_{xz} & E_{xy} + E_{yx} & -E_{xx} + E_{yy} - E_{zz} & E_{yz} + E_{zy} \\ E_{xy} - E_{yx} & E_{zx} + E_{xz} & E_{yz} + E_{zy} & -E_{xx} - E_{yy} + E_{zz} \end{bmatrix} \quad (13)$$

built from our original 3×3 cross-covariance matrix E defined by equation (11). We will refer to $M(E)$ from here on as the *profile matrix*, as it essentially reveals a different viewpoint of the optimization function and its relationship to

the matrix E . Note that in some literature matrices related to the cross-covariance matrix E may be referred to as ‘attitude profile matrices’ and one also may see the term ‘key matrix’ referring to $M(E)$.

The bottom line is that if one decomposes equation (13) into its eigensystem, the measure equation (12) is maximized when the unit-length quaternion vector q is the eigenvector of $M(E)$ ’s largest eigenvalue (Davenport, 1968; Faugeras & Hebert, 1983; Horn, 1987; Diamond, 1988; Kearsley, 1989; Kneller, 1991). The RMSD optimal-rotation problem thus reduces to finding the maximal eigenvalue ϵ_{opt} of $M(E)$ (which we emphasize depends only on the numerical data). Plugging the corresponding eigenvector q_{opt} into equation (5), we obtain the rotation matrix $R(q_{\text{opt}})$ that solves the problem. The resulting proximity measure relating $\{x_k\}$ and $\{y_k\}$ is simply

$$\left. \begin{aligned} \Delta_{\text{opt}} &= q_{\text{opt}} \cdot M(E) \cdot q_{\text{opt}} \\ &= q_{\text{opt}} \cdot (\epsilon_{\text{opt}} q_{\text{opt}}) \\ &= \epsilon_{\text{opt}} \end{aligned} \right\} \quad (14)$$

and does not require us to actually compute q_{opt} or $R(q_{\text{opt}})$ explicitly if all we want to do is compare various test data sets to a reference structure.

Note. In the interests of conceptual and notational simplicity, we have made a number of assumptions. For one thing, in declaring that equation (5) describes our sought-for rotation matrix, we have presumed that the optimal rotation matrix will always be a proper rotation, with $\det R = +1$. Also, as mentioned, we have omitted any general translation problems, assuming that there is a way to translate each data set to an appropriate center, e.g. by subtracting the center of mass. The global translation optimization process is treated in Faugeras & Hebert (1986) and Coutsias *et al.* (2004), and discussions of center-of-mass alignment, scaling and point weighting are given in much of the original literature, see, e.g., Horn (1987), Coutsias *et al.* (2004), and Theobald (2005). Finally, in real problems, structures such as molecules may appear in mirror-image or enantiomer form, and such issues were introduced early on by Kabsch (1976, 1978). There can also be particular symmetries, or very close approximations to symmetries, that can make some of our natural assumptions about the good behavior of the profile matrix invalid, and many of these issues, including ways to treat degenerate cases, have been carefully studied [see, e.g., Coutsias *et al.* (2004) and Coutsias & Wester (2019)]. The latter authors also point out that if a particular data set $M(E)$ produces a negative smallest eigenvalue ϵ_4 such that $|\epsilon_4| > \epsilon_{\text{opt}}$, this can be a sign of a reflected match, and the *negative* rotation matrix $R_{\text{opt}} = -R(q(\epsilon_4))$ may actually produce the best alignment. These considerations may be essential in some applications, and readers are referred to the original literature for details.

5.4. Illustrative example

We can visualize the transition from the initial data $\Delta(q_{\text{ID}}) = \text{tr } E$ to the optimal alignment $\Delta(q_{\text{opt}}) = \epsilon_{\text{opt}}$ by exploiting the geodesic interpolation equation (8) from the identity quaternion q_{ID} to q_{opt} given by

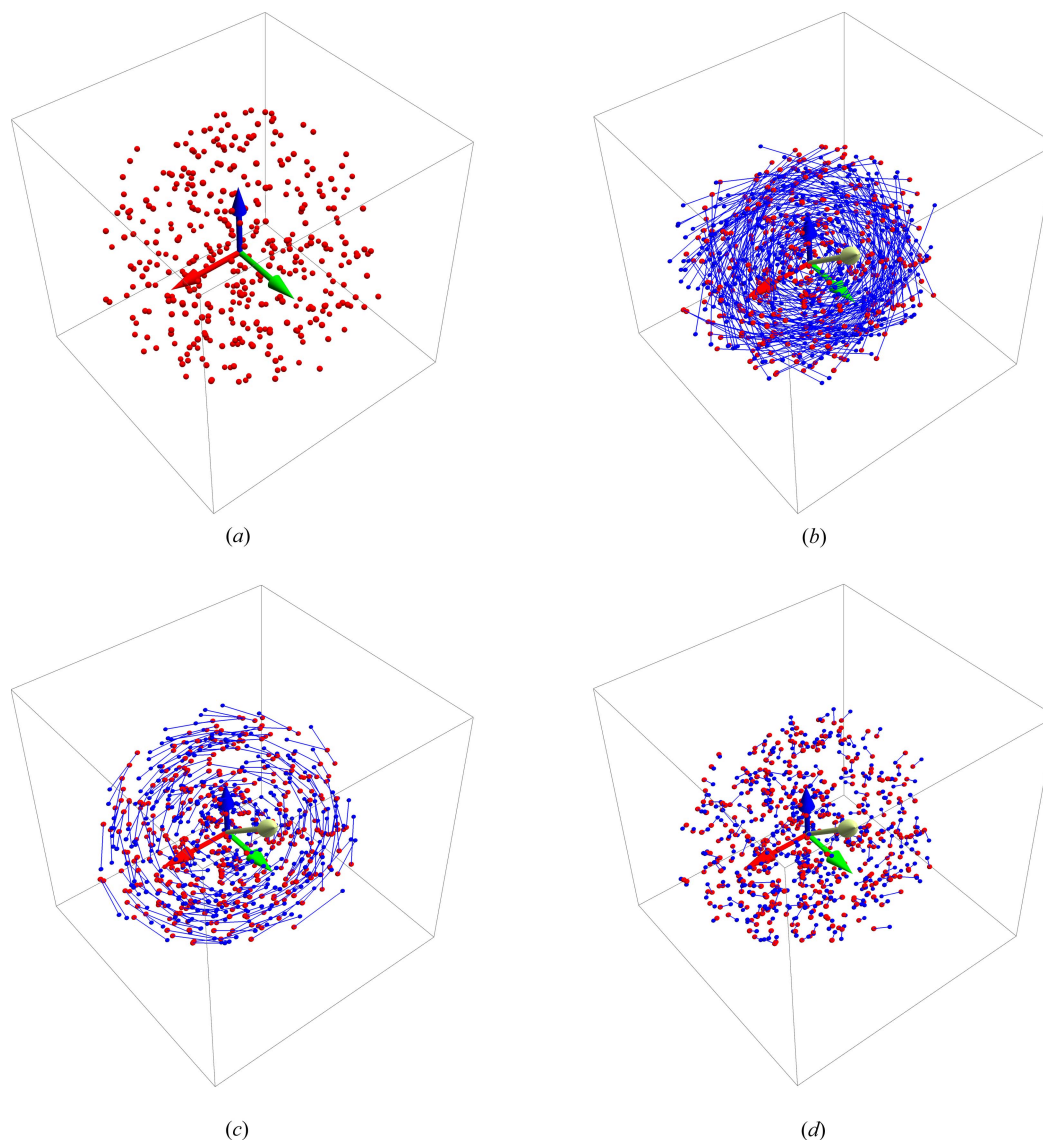


Figure 2

(a) A typical 3D spatial reference data set. (b) The reference data in red alongside the test data in blue, with blue lines representing the Euclidean distances connecting each test data point with its corresponding reference point. (c) The partial alignment at $s = 0.75$. (d) The optimal alignment for this data set at $s = 1.0$. The yellow arrow is the axis of rotation specified by the optimal quaternion's spatial components.

$$q(s) = \text{slerp}(q_{\text{ID}}, q_{\text{opt}}, s) \quad (15)$$

and applying the resulting rotation matrix $R(q(s))$ to the test data, ending with $R(q_{\text{opt}})$ showing the best alignment of the two data sets. In Fig. 2, we show a sample reference data set in red, a sample test data set in blue connected to the reference data set by blue lines, an intermediate partial alignment and finally the optimally aligned pair. The yellow arrow is the *spatial part of the quaternion solution*, proportional to the eigenvector $\hat{\mathbf{n}}$ (fixed axis) of the optimal 3D rotation matrix $R(q) = R(\theta, \hat{\mathbf{n}})$, and whose length is $\sin(\theta/2)$, sine of half the rotation angle needed to perform the optimal alignment of the test data with the reference data. In Fig. 3, we visualize the optimization process in an alternative way, showing random samples of $q = (q_0, \mathbf{q})$ in \mathbf{S}^3 , separated into the 'northern hemisphere' 3D unit-radius ball in (a) with $q_0 \geq 0$, and the 'southern hemisphere' 3D unit-

radius ball in (b) with $q_0 \leq 0$. (This is like representing the Earth as two flattened discs, one showing everything above the equator and one showing everything below the equator; the distance from the equatorial plane is implied by the location in the disc, with the maximum at the centers, the north and south poles.) Either solid ball contains one unique quaternion for every possible choice of $R(q)$, modulo the doubling of diametrically opposite points at $q_0 = 0$. The values of $\Delta(q) = \text{tr} R(q) \cdot E$ are shown as scaled dots located at their corresponding spatial ('imaginary') quaternion points \mathbf{q} in the solid balls. The yellow arrows, equivalent negatives of each other, show the spatial part \mathbf{q}_{opt} of the optimal quaternion q_{opt} , and the tips of the arrows clearly fall in the middle of the mirror pair of clusters of the largest values of $\Delta(q)$. Note that the lower-left dots in (a) continue smoothly into the larger lower-left dots in (b), which is the center of the optimal quaternion in (b). Further details of such methods of

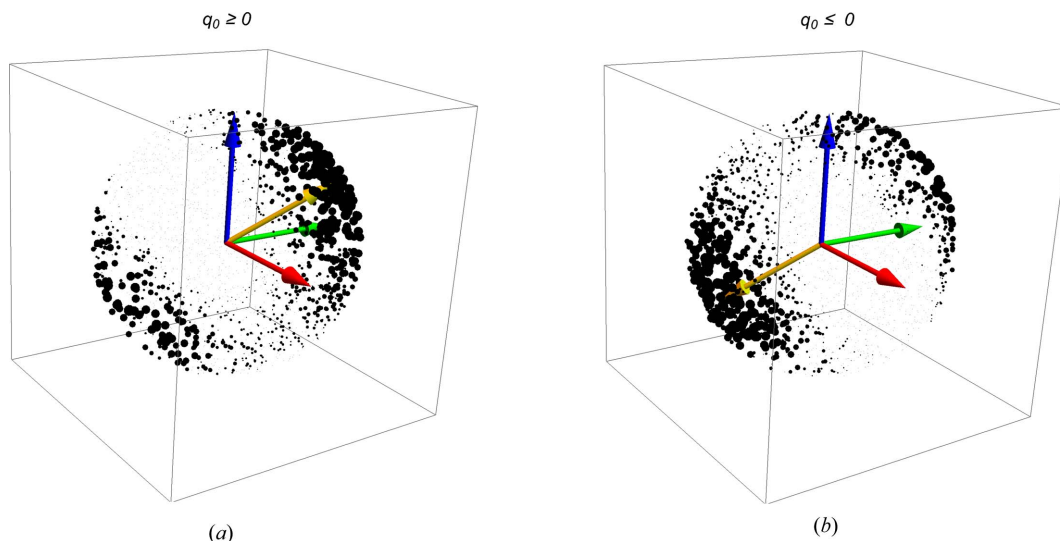


Figure 3

The values of $\Delta(q) = \text{tr} R(q) \cdot E = q \cdot M(E) \cdot q$ represented by the sizes of the dots placed randomly in the ‘northern’ and ‘southern’ 3D solid balls spanning the entire hypersphere S^3 with (a) containing the $q_0 \geq 0$ sector and (b) containing the $q_0 \leq 0$ sector. We display the data dots at the locations of their spatial quaternion components $\mathbf{q} = (q_1, q_2, q_3)$, and we know that $q_0 = \pm(1 - \mathbf{q} \cdot \mathbf{q})^{1/2}$ so the \mathbf{q} data uniquely specify the full quaternion. Since $R(q) = R(-q)$, the points in *each* ball actually represent all possible unique rotation matrices. The spatial component of the maximal eigenvector is shown by the yellow arrows, which clearly end in the middle of the maximum values of $\Delta(q)$. Note that, in the quaternion context, diametrically opposite points on the spherical surface are identical rotations, so the cluster of larger dots at the upper right of (a) is, in the entire sphere, representing the same data as the ‘diametrically opposite’ lower-left cluster in (b), both surrounding the tips of their own yellow arrows. The smaller dots at the upper right of (b) are contiguous with the upper-right region of (a), forming a single cloud centered on \mathbf{q}_{opt} , and similarly for the lower left of (a) and the lower left of (b). The whole figure contains two distinct clusters of dots (related by $q \rightarrow -q$) centered around $\pm\mathbf{q}_{\text{opt}}$.

displaying quaternions are provided in Appendix B [see also Hanson (2006)].

6. Algebraic solution of the eigensystem for 3D spatial alignment

At this point, one can simply use the traditional *numerical* methods to solve equation (12) for the maximal eigenvalue ϵ_{opt} of $M(E)$ and its eigenvector q_{opt} , thus solving the 3D spatial-alignment problem of equation (10). Alternatively, we can also exploit *symbolic* methods to study the properties of the eigensystems of 4×4 matrices M algebraically to provide deeper insights into the structure of the problem, and that is the subject of this section.

Theoretically, the algebraic form of our eigensystem is a textbook problem following from the 16th-century-era solution of the quartic algebraic equation in, *e.g.*, Abramowitz & Stegun (1970). Our objective here is to explore this textbook solution in the specific context of its application to eigensystems of 4×4 matrices and its behavior relative to the properties of such matrices. The real, symmetric, traceless profile matrix $M(E)$ in equation (13) appearing in the 3D spatial RMSD optimization problem must necessarily possess only real eigenvalues, and the properties of $M(E)$ permit some particular simplifications in the algebraic solutions that we will discuss. The quaternion RMSD literature varies widely in the details of its treatment of the algebraic solutions, ranging from no discussion at all, to Horn, who mentions the possibility but does not explore it, to the work of Coutsiias *et al.* (Coutsiias *et al.*, 2004; Coutsiias & Wester, 2019), who present an exhaustive treatment, in addition to working out the exact details of the

correspondence between the SVD eigensystem and the quaternion eigensystem, both of which in principle embody the algebraic solution to the RMSD optimization problem. In addition to the treatment of Coutsiias *et al.*, other approaches similar to the one we will study are due to Euler (Euler, 1733; Bell, 2008), as well as a series of papers on the quartic by Nickalls (1993, 2009).

6.1. Eigenvalue expressions

We begin by writing down the eigenvalue expansion of the profile matrix,

$$\det[M - eI_4] = e^4 + e^3 p_1 + e^2 p_2 + e p_3 + p_4 = 0, \quad (16)$$

where e denotes a generic eigenvalue, I_4 is the 4D identity matrix and the p_k are homogeneous polynomials of degree k in the elements of M . For the special case of a traceless, symmetric profile matrix $M(E)$ defined by equation (13), the $p_k(E)$ coefficients simplify and can be expressed numerically as the following functions either of M or of E :

$$p_1(E) = -\text{tr}[M] = 0 \quad (17)$$

$$\begin{aligned} p_2(E) &= -\frac{1}{2} \text{tr}[M \cdot M] = -2 \text{tr}[E \cdot E^T] \\ &= -2(E_{xx}^2 + E_{xy}^2 + E_{xz}^2 + E_{yx}^2 + E_{yy}^2 + E_{yz}^2 \\ &\quad + E_{zx}^2 + E_{zy}^2 + E_{zz}^2) \end{aligned} \quad (18)$$

$$\begin{aligned} p_3(E) &= -\frac{1}{3} \text{tr}[M \cdot M \cdot M] = -8 \det[E] \\ &= 8(E_{xx} E_{yz} E_{zy} + E_{yy} E_{xz} E_{zx} + E_{zz} E_{xy} E_{yx}) \\ &\quad - 8(E_{xx} E_{yy} E_{zz} + E_{xy} E_{yz} E_{zx} + E_{xz} E_{zy} E_{yx}) \end{aligned} \quad (19)$$

$$p_4(E) = \det[M] = 2 \text{tr}[E \cdot E^T \cdot E \cdot E^T] - (\text{tr}[E \cdot E^T])^2. \quad (20)$$

Interestingly, the polynomial $M(E)$ is arranged so that $-p_2(E)/2$ is the (squared) Fröbenius norm of E , and $-p_3(E)/8$ is its determinant. Our task now is to express the four eigenvalues $e = \epsilon_k(p_1, p_2, p_3, p_4)$, $k = 1, \dots, 4$, usefully in terms of the matrix elements, and also to find their eigenvectors; we are of course particularly interested in the maximal eigenvalue ϵ_{opt} .

6.2. Approaches to algebraic solutions

Equation (16) can be solved directly using the quartic equations published by Cardano in 1545 [see, e.g., Abramowitz & Stegun (1970), Weisstein (2019b), and Wikipedia (2019)], which are incorporated into the *Mathematica* function

$$\text{Solve}[\text{myQuarticEqn}[e] == 0, e, \text{Quartics} \rightarrow \text{True}] \quad (21)$$

that immediately returns a suitable algebraic formula. At this point we defer detailed discussion of the textbook solution to the supporting information, and instead focus on a particularly symmetric version of the solution and the form it takes for the eigenvalue problem for traceless, symmetric 4×4 matrices such as our profile matrices $M(E)$. For this purpose, we look for an alternative solution by considering the following traceless ($p_1 = 0$) ansatz:

$$\left. \begin{aligned} \epsilon_1(p) &\stackrel{?}{=} \sqrt{X(p)} + \sqrt{Y(p)} + \sqrt{Z(p)} \\ \epsilon_2(p) &\stackrel{?}{=} \sqrt{X(p)} - \sqrt{Y(p)} - \sqrt{Z(p)} \\ \epsilon_3(p) &\stackrel{?}{=} \sqrt{X(p)} + \sqrt{Y(p)} - \sqrt{Z(p)} \\ \epsilon_4(p) &\stackrel{?}{=} \sqrt{X(p)} - \sqrt{Y(p)} + \sqrt{Z(p)} \end{aligned} \right\}. \quad (22)$$

This form emphasizes some additional explicit symmetry that we will see is connected to the role of cube roots in the quartic algebraic solutions (see, e.g., Coutsias & Wester, 2019). We can turn it into an equation for $\epsilon_k(p)$ to be solved in terms of the matrix parameters $p_k(E)$ as follows: First we eliminate e using $(e - \epsilon_1)(e - \epsilon_2)(e - \epsilon_3)(e - \epsilon_4) = 0$ to express the matrix data expressions p_k directly in terms of totally symmetric polynomials of the eigenvalues in the form (Abramowitz & Stegun, 1970)

$$\left. \begin{aligned} p_1 &= -\epsilon_1 - \epsilon_2 - \epsilon_3 - \epsilon_4 \\ p_2 &= \epsilon_1\epsilon_2 + \epsilon_1\epsilon_3 + \epsilon_2\epsilon_3 + \epsilon_1\epsilon_4 + \epsilon_2\epsilon_4 + \epsilon_3\epsilon_4 \\ p_3 &= -\epsilon_1\epsilon_2\epsilon_3 - \epsilon_1\epsilon_2\epsilon_4 - \epsilon_1\epsilon_3\epsilon_4 - \epsilon_2\epsilon_3\epsilon_4 \\ p_4 &= \epsilon_1\epsilon_2\epsilon_3\epsilon_4 \end{aligned} \right\}. \quad (23)$$

Next we substitute our expression equation (22) for the ϵ_k in terms of the $\{X, Y, Z\}$ functions into equation (23), yielding a completely different alternative to equation (16) that will *also* solve the 3D RMSD eigenvalue problem if we can invert it to express $\{X(p), Y(p), Z(p)\}$ in terms of the data $p_k(E)$ as presented in equation (20):

$$\left. \begin{aligned} p_1 &= 0 \\ p_2 &= -2(X + Y + Z) \\ p_3 &= -8\sqrt{XYZ} \\ p_4 &= X^2 + Y^2 + Z^2 - 2(YZ + ZX + XY) \end{aligned} \right\}. \quad (24)$$

We already see the critical property in p_3 that, while p_3 itself has a deterministic sign from the matrix data, the possibly variable signs of the square roots in equation (22) have to be constrained so their product \sqrt{XYZ} agrees with the sign of p_3 . Manipulating the quartic equation solutions that we can obtain by applying the library function equation (21) to equation (24), and restricting our domain to real traceless, symmetric matrices (and hence real eigenvalues), we find solutions for $X(p)$, $Y(p)$ and $Z(p)$ of the following form:

$$F_f(0, p_2, p_3, p_4) = +\frac{1}{6}(r(p) \cos_f(p) - p_2), \quad (25)$$

where the $\cos_f(p)$ terms differ only by a cube-root phase:

$$\left. \begin{aligned} \cos_x(p) &= \cos\left(\frac{\arg(a + ib)}{3}\right), \\ \cos_y(p) &= \cos\left(\frac{\arg(a + ib)}{3} - \frac{2\pi}{3}\right), \\ \cos_z(p) &= \cos\left(\frac{\arg(a + ib)}{3} + \frac{2\pi}{3}\right). \end{aligned} \right\}. \quad (26)$$

Here $\arg(a + ib) = \text{atan2}(b, a)$ in the C mathematics library, or $\text{ArcTan}[a, b]$ in *Mathematica*, $F_f(p)$ with $f = (x, y, z)$ corresponds to $X(p)$, $Y(p)$ or $Z(p)$, and the utility functions appearing in the equations for our traceless $p_1 = 0$ case are

$$\left. \begin{aligned} r^2(0, p_2, p_3, p_4) &= p_2^2 + 12p_4 = (a^2 + b^2)^{1/3} \\ &= (a + ib)^{1/3}(a - ib)^{1/3} \\ a(0, p_2, p_3, p_4) &= p_2^3 + \frac{1}{2}(27p_3^2 - 72p_2p_4) \\ b^2(0, p_2, p_3, p_4) &= r^6(p) - a^2(p) \\ &= \frac{27}{4}(16p_4p_2^4 - 4p_3^2p_2^3 - 128p_4^2p_2^2 \\ &\quad + 144p_3^2p_4p_2 - 27p_3^4 + 256p_4^3) \end{aligned} \right\}. \quad (27)$$

The function $b^2(p)$ has the essential property that, for real solutions to the cubic, which imply the required real solutions to our eigenvalue equations (Abramowitz & Stegun, 1970), we must have $b^2(p) \geq 0$. That essential property allowed us to convert the bare solution into terms involving $\{(a + ib)^{1/3}, (a - ib)^{1/3}\}$ whose sums form the manifestly real cube-root-related cosine terms in equation (26).

6.3. Final eigenvalue algorithm

While equations (25) and (26) are well defined, square roots must be taken to finish the computation of the eigenvalues postulated in equation (22). In our special case of symmetric, traceless matrices such as $M(E)$, we can always choose the signs of the first two square roots to be positive, but the sign of the \sqrt{Z} term is non-trivial, and in fact is the sign of $\det[E]$. The form of the solution in equations (22) and (25) that works specifically for all traceless symmetric matrices such as $M(E)$ is given by our equations for $p_k(E)$ in equations (17)–(22), along with equations (25), (26) and (27) *provided* we modify equation (22) using $\sigma(p) = \text{sign}(\det[E]) = \text{sign}(-p_3)$ as follows:

$$\left. \begin{aligned} \epsilon_1(p) &= \sqrt{X(p)} + \sqrt{Y(p)} + \sigma(p)\sqrt{Z(p)} \\ \epsilon_2(p) &= \sqrt{X(p)} - \sqrt{Y(p)} - \sigma(p)\sqrt{Z(p)} \\ \epsilon_3(p) &= \sqrt{X(p)} + \sqrt{Y(p)} - \sigma(p)\sqrt{Z(p)} \\ \epsilon_4(p) &= \sqrt{X(p)} - \sqrt{Y(p)} + \sigma(p)\sqrt{Z(p)} \end{aligned} \right\}. \quad (28)$$

The particular order of the numerical eigenvalues in our chosen form of the solution equation (28) is found in regular cases to be uniformly non-increasing in numerical order for our $M(E)$ matrices, so $\epsilon_1(p)$ is always the leading eigenvalue. This is our preferred symbolic version of the solution to the 3D RMSD problem defined by $M(E)$.

Note: We have experimentally confirmed the numerical behavior of equation (25) in equation (28) with 1 000 000 randomly generated sets of 3D cross-covariance matrices E , along with the corresponding profile matrices $M(E)$, producing numerical values of p_k inserted into the equations for $X(p)$, $Y(p)$ and $Z(p)$. We confirmed that the sign of $\sigma(p)$ varied randomly, and found that the algebraically computed values of $\epsilon_k(p)$ corresponded to the standard numerical eigenvalues of the matrices $M(E)$ in all cases, to within expected variations due to numerical evaluation behavior and expected occasional instabilities. In particular, we found a maximum per-eigenvalue discrepancy of about 10^{-13} for the *algebraic* methods relative to the standard *numerical* eigenvalue methods, and a median difference of 10^{-15} , in the context of machine precision of about 10^{-16} . (Why did we do this? Because we had earlier versions of the algebraic formulas that produced anomalies due to inconsistent phase choices in the roots, and felt it worthwhile to perform a practical check on the numerical behavior of our final version of the solutions.)

6.4. Eigenvectors for 3D data

The eigenvector formulas corresponding to ϵ_k can be generically computed by solving any three rows of

$$[M(E) \cdot v - e v] = [A] = 0 \quad (29)$$

for the elements of v , e.g. $v = (1, v_1, v_2, v_3)$, as a function of some eigenvalue e (of course, one must account for special cases, e.g. if some subspace of $M(E)$ is already diagonal). The desired unit quaternion for the optimization problem can then be obtained from the normalized eigenvector

$$q(e, E) = \frac{v}{\|v\|}. \quad (30)$$

Note that this can often have $q_0 < 0$, and that whenever the problem in question depends on the sign of q_0 , such as a slerp starting at q_{ID} , one should choose the sign of equation (30) appropriately; some applications may also require an element of statistical randomness, in which case one might randomly pick a sign for q_0 .

As noted by Liu *et al.* (2010), a very clear way of computing the eigenvectors for a given eigenvalue is to exploit the fact that the determinant of equation (29) must vanish, that is $\det[A] = 0$; one simply exploits the fact that the columns of the adjugate matrix α_{ij} (the transpose of the matrix of cofac-

tors of the matrix $[A]$) produce its inverse by means of creating multiple copies of the determinant. That is,

$$\sum_{c=1}^4 A_{ac} \alpha_{cb} = \delta_{ab} \det[A] \equiv 0, \quad (31)$$

so we can just compute *any* column of the adjugate via the appropriate set of subdeterminants and, in the absence of singularities, that will be an eigenvector (since any of the four columns can be eigenvectors, if one fails just try another).

In the general well behaved case, the form of v in the eigenvector solution for any eigenvalue $e = \epsilon_k$ may be explicitly computed to give the corresponding quaternion (among several equivalent alternative expressions) as

$$q(e, E) = \frac{1}{\|v\|} \begin{bmatrix} 2ABC + A^2e_x + B^2e_y + C^2e_z - e_x e_y e_z \\ A(aA - bB - cC) - cBe_y - bCe_z - a e_y e_z \\ B(bB - cC - aA) - aCe_z - cAe_x - b e_z e_x \\ C(cC - aA - bB) - bAe_x - aBe_y - c e_x e_y \end{bmatrix}, \quad (32)$$

where for convenience we define $\{e_x = (e - x + y + z), e_y = (e + x - y + z), e_z = (e + x + y - z)\}$ with $x = E_{xx}$, cyclic, $a = E_{yz} - E_{zy}$, cyclic, and $A = E_{yz} + E_{zy}$, cyclic. We substitute the maximal eigenvector $q_{opt} = q(\epsilon_1, E)$ into equation (5) to give the sought-for optimal 3D rotation matrix $R(q_{opt})$ that solves the RMSD problem with $\Delta(q_{opt}) = \epsilon_1$, as we noted in equation (14).

Remark: Yet another approach to computing eigenvectors that, surprisingly, *almost* entirely avoids any reference to the original matrix, but needs only its eigenvalues and minor eigenvalues, has recently been rescued from relative obscurity (Denton *et al.*, 2019). (The authors uncovered a long list of non-cross-citing literature mentioning the result dating back at least to 1934.) If, for a real, symmetric 4×4 matrix M we label the set of four eigenvectors v_i by the index i and the components of any single such four-vector by a , the squares of each of the sixteen corresponding components take the form

$$([v_i]_a)^2 = \frac{\prod_{j=1}^3 (\lambda_j(M) - \lambda_j(\mu_a))}{\prod_{k=1; k \neq i}^4 (\lambda_i(M) - \lambda_k(M))}. \quad (33)$$

Here the μ_a are the 3×3 minors obtained by removing the a th row and column of M , and the $\lambda_j(\mu_a)$ comprise the list of three eigenvalues of each of these minors. Attempting to obtain the eigenvectors by taking square roots is of course hampered by the nondeterministic sign; however, since the eigenvalues $\lambda_i(M)$ are known, and the overall sign of each eigenvector v_i is arbitrary, one needs to check at most eight sign combinations to find the one for which $M \cdot v_i = \lambda_i(M)v_i$, solving the problem. Note that the general formula extends to Hermitian matrices of any dimension.

7. The 3D orientation-frame alignment problem

We turn next to the orientation-frame problem, assuming that the data are like lists of orientations of rollercoaster cars,

or lists of residue orientations in a protein, ordered pairwise in some way, but *without* specifically considering any spatial location or nearest-neighbor ordering information. In D -dimensional space, the *columns* of any $\mathbf{SO}(D)$ orthonormal $D \times D$ rotation matrix R_D are what we mean by an orientation frame, since these columns are the directions pointed to by the axes of the identity matrix after rotating something from its defining identity frame to a new attitude; note that no spatial location information whatever is contained in R_D , though one may wish to choose a local center for each frame if the construction involves coordinates such as amino-acid atom locations (see, *e.g.*, Hanson & Thakur, 2012).

In 2D, 3D and 4D, there exist two-to-one quadratic maps from the topological spaces \mathbf{S}^1 , \mathbf{S}^3 and $\mathbf{S}^3 \times \mathbf{S}^3$ to the rotation matrices R_2 , R_3 and R_4 . These are the quaternion-related objects that we will use to obtain elegant representations of the frame data-alignment problem. In 2D, a frame data element can be expressed as a complex phase, while in 3D the frame is a unit quaternion [see Hanson (2006) and Hanson & Thakur (2012)]. In 4D (see the supporting information), the frame is described by a *pair* of unit quaternions.

Note. Readers unfamiliar with the use of complex numbers and quaternions to obtain elegant representations of 2D and 3D orientation frames are encouraged to review the tutorial in Appendix B.

7.1. Overview

We focus now on the problem of aligning corresponding sets of 3D orientation frames, just as we already studied the alignment of sets of 3D spatial coordinates by performing an optimal rotation. There will be more than one feasible method. We might assume we could just define the quaternion-frame alignment or ‘QFA’ problem by converting any list of frame orientation matrices to quaternions [see Hanson (2006), Hanson & Thakur (2012) and also Appendix C] and writing down the quaternion equivalents of the RMSD treatment in equation (9) and equation (10). However, unlike the linear Euclidean problem, the preferred quaternion optimization function technically requires a *nonlinear* minimization of the squared sums of geodesic arc lengths connecting the points on the quaternion hypersphere \mathbf{S}^3 . The task of formulating this ideal problem as well as studying alternative approximations is the subject of its own branch of the literature, often known as the *quaternionic barycenter* problem or the *quaternion averaging* problem (see, *e.g.*, Brown & Worsey, 1992; Buss & Fillmore, 2001; Moakher, 2002; Markley *et al.*, 2007; Huynh, 2009; Hartley *et al.*, 2013; and also Appendix D). We will focus on L_2 norms (the aforementioned sums of squares of arc lengths), although alternative approaches to the rotation-averaging problem, such as employing L_1 norms and using the Weiszfeld algorithm to find the optimal rotation numerically, have been advocated, *e.g.*, by Hartley *et al.* (2011). The computation of optimally aligning rotations, based on plausible exact or approximate measures relating collections of corresponding pairs of (quaternionic) orientation frames, is now our task.

Choices for the forms of the measures encoding the distance between orientation frames have been widely discussed, see, *e.g.*, Park & Ravani (1997), Moakher (2002), Markley *et al.* (2007), Huynh (2009), Hartley *et al.* (2011, 2013), and Huggins (2014a). Since we are dealing primarily with quaternions, we will start with two measures dealing directly with the quaternion geometry, the geodesic arc length and the chord length, and later on examine some advantages of starting with quaternion-sign-independent rotation-matrix forms.

7.2. 3D geodesic arc-length distance

First, we recall that the matrix equation (5) has three orthonormal columns that define a quadratic map from the quaternion three-sphere \mathbf{S}^3 , a smooth connected Riemannian manifold, to a 3D orientation frame. The squared geodesic arc-length distance between two quaternions lying on the three-sphere \mathbf{S}^3 is generally agreed upon as the measure of orientation-frame proximity whose properties are the closest in principle to the ordinary squared Euclidean distance measure equation (9) between points (Huynh, 2009), and we will adopt this measure as our starting point. We begin by writing down a frame–frame distance measure between two unit quaternions q_1 and q_2 , corresponding precisely to two orientation frames defined by the columns of $R(q_1)$ and $R(q_2)$. We define the geodesic arc length as an angle α on the hypersphere \mathbf{S}^3 computed geometrically from $q_1 \cdot q_2 = \cos \alpha$. As pointed out by Huynh (2009) and Hartley *et al.* (2013), the geodesic arc length between a test quaternion q_1 and a data-point quaternion q_2 of ambiguous sign [since $R(+q_2) = R(-q_2)$] can take two values, and we want the minimum value. Furthermore, to work on a spherical manifold instead of a plane, we need basically to cluster the ambiguous points in a deterministic way. Starting with the bare angle between two quaternions on \mathbf{S}^3 , $\alpha = \arccos(q_1 \cdot q_2)$, where we recall that $\alpha \geq 0$, we define a *pseudometric* (Huynh, 2009) for the geodesic arc-length distance as

$$d_{\text{geodesic}}(q_1, q_2) = \min(\alpha, \pi - \alpha): \quad 0 \leq d_{\text{geodesic}}(q_1, q_2) \leq \frac{\pi}{2}, \quad (34)$$

as illustrated in Fig. 4. An efficient implementation of this is to take

$$d_{\text{geodesic}}(q_1, q_2) = \arccos(|q_1 \cdot q_2|). \quad (35)$$

We now seek to define an ideal minimizing L_2 orientation-frame measure, comparable to our minimizing Euclidean RMSD measure, but constructed from geodesic arc lengths on the quaternion hypersphere instead of Euclidean distances in space. Thus to compare a test quaternion-frame data set $\{p_k\}$ to a reference data set $\{r_k\}$, we propose the geodesic based least-squares measure

$$\begin{aligned} \mathbf{S}_{\text{geodesic}} &= \sum_{k=1}^N (\arccos |(q \star p_k) \cdot r_k|)^2 \\ &= \sum_{k=1}^N (\arccos |q \cdot (r_k \star \bar{p}_k)|)^2, \end{aligned} \quad (36)$$

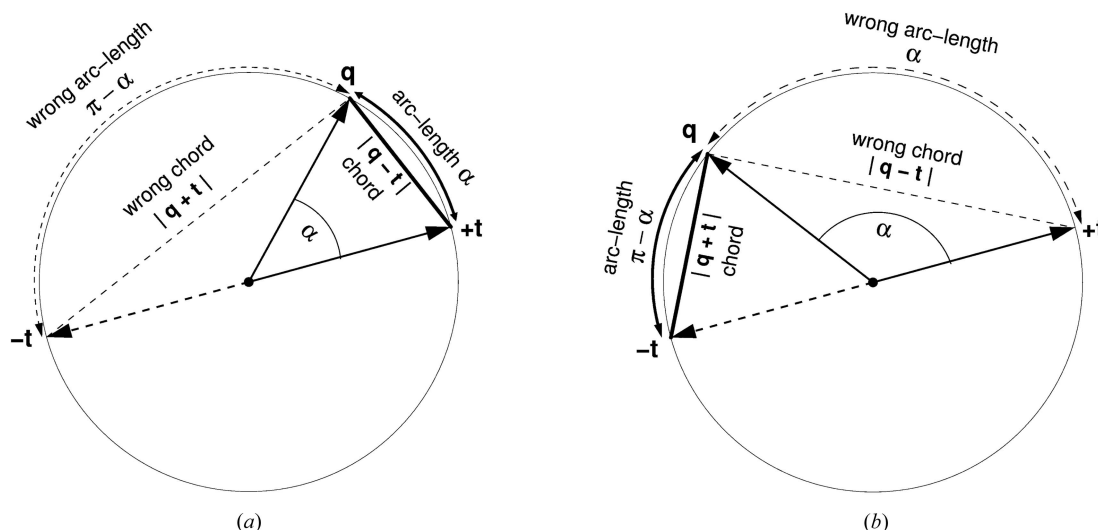


Figure 4 Geometric context involved in choosing a *quaternion distance* that will result in the correct *average rotation matrix* when the quaternion measures are optimized. Because the quaternion vectors represented by t and $-t$ give the same rotation matrix, one must choose $|\cos \alpha|$ or the *minima*, that is $\min(\alpha, \pi - \alpha)$ or $\min(\|q - t\|, \|q + t\|)$, of the alternative distance measures to get the *correct* items in the arc-length or chord measure summations. (a) and (b) represent the cases when the first or second choice should be made, respectively.

where we have used the identities of equation (3). When $q = q_{ID}$, the individual measures correspond to equation (35), and otherwise ‘ $q \star p_k$ ’ is the exact analog of ‘ $R(q) \cdot x_k$ ’ in equation (9), and denotes the quaternion rotation q acting on the entire set $\{p_k\}$ to rotate it to a new orientation that we want to align optimally with the reference frames $\{r_k\}$. Analogously, for points on a sphere, the arccosine of an inner product is equivalent to a distance between points in Euclidean space.

Remark: For improved numerical behavior in the computation of the quaternion inner-product angle between two quaternions, one may prefer to convert the arccosine to an arctangent form, $\alpha = \arctan(dx, dy) = \arctan(\cos \alpha, |\sin \alpha|)$ [remember the C math library uses the opposite argument order $\text{atan2}(dy, dx)$], with the parameters

$$\begin{aligned} \cos(\alpha) &= |\Re(q_1 \star q_2^{-1})| = |q_1 \cdot q_2|, \\ |\sin(\alpha)| &= \|\Im(q_1 \star q_2^{-1})\| = \|-[q_1]_0 \mathbf{q}_2 + [q_2]_0 \mathbf{q}_1 - \mathbf{q}_1 \times \mathbf{q}_2\|, \end{aligned}$$

which is somewhat more stable.

7.3. Adopting the solvable chord measure

Unfortunately, the geodesic arc-length measure does not fit into the linear algebra approach that we were able to use to obtain exact solutions for the Euclidean-data-alignment problem treated so far. Thus we are led to investigate instead a very close approximation to $d_{\text{geodesic}}(q_1, q_2)$ that *does* correspond closely to the Euclidean data case and does, with some contingencies, admit exact solutions. This approximate measure is the *chord distance*, whose individual distance terms analogous to equation (35) take the form of a closely related pseudometric (Huynh, 2009; Hartley *et al.*, 2013),

$$\begin{aligned} d_{\text{chord}}(q_1, q_2) &= \min(\|q_1 - q_2\|, \|q_1 + q_2\|): \\ 0 &\leq d_{\text{chord}}(q_1, q_2) \leq \sqrt{2}. \end{aligned} \quad (37)$$

We compare the geometric origins for equation (35) and equation (37) in Fig. 4. Note that the crossover point between the two expressions in equation (37) is at $\pi/2$, so the hypotenuse of the right isosceles triangle at that point has length $\sqrt{2}$.

The solvable approximate optimization function analogous to $\|R \cdot x - y\|^2$ that we will now explore for the quaternion-frame alignment problem will thus take the form that must be minimized as

$$S_{\text{chord}} = \sum_{k=1}^N (\min(\|(q \star p_k) - r_k\|, \|(q \star p_k) + r_k\|))^2. \quad (38)$$

We can convert the sign ambiguity in equation (38) to a deterministic form like equation (35) by observing, with the help of Fig. 4, that

$$\|q_1 - q_2\|^2 = 2 - 2q_1 \cdot q_2, \quad \|q_1 + q_2\|^2 = 2 + 2q_1 \cdot q_2. \quad (39)$$

Clearly $(2 - 2|q_1 \cdot q_2|)$ is always the smallest of the two values. Thus minimizing equation (38) amounts to maximizing the now-familiar cross-term form, which we can write as

$$\begin{aligned} \Delta_{\text{chord}}(q) &= \sum_{k=1}^N |(q \star p_k) \cdot r_k| \\ &= \sum_{k=1}^N |q \cdot (r_k \star \bar{p}_k)| \\ &= \sum_{k=1}^N |q \cdot t_k| \end{aligned} \quad (40)$$

Here we have used the identity $(q \star p) \cdot r = q \cdot (r \star \bar{p})$ from equation (3) and defined the quaternion displacement or ‘attitude error’ (Markley *et al.*, 2007)

$$t_k = r_k \star \bar{p}_k. \quad (41)$$

Note that we could have derived the same result using equation (2) to show that $\|q \star p - r\| = \|q \star p - r\| \|p\| = \|q - r \star \bar{p}\|$.

There are several ways to proceed to our final result at this point. The simplest is to pick a neighborhood in which we will choose the samples of q that include our expected optimal

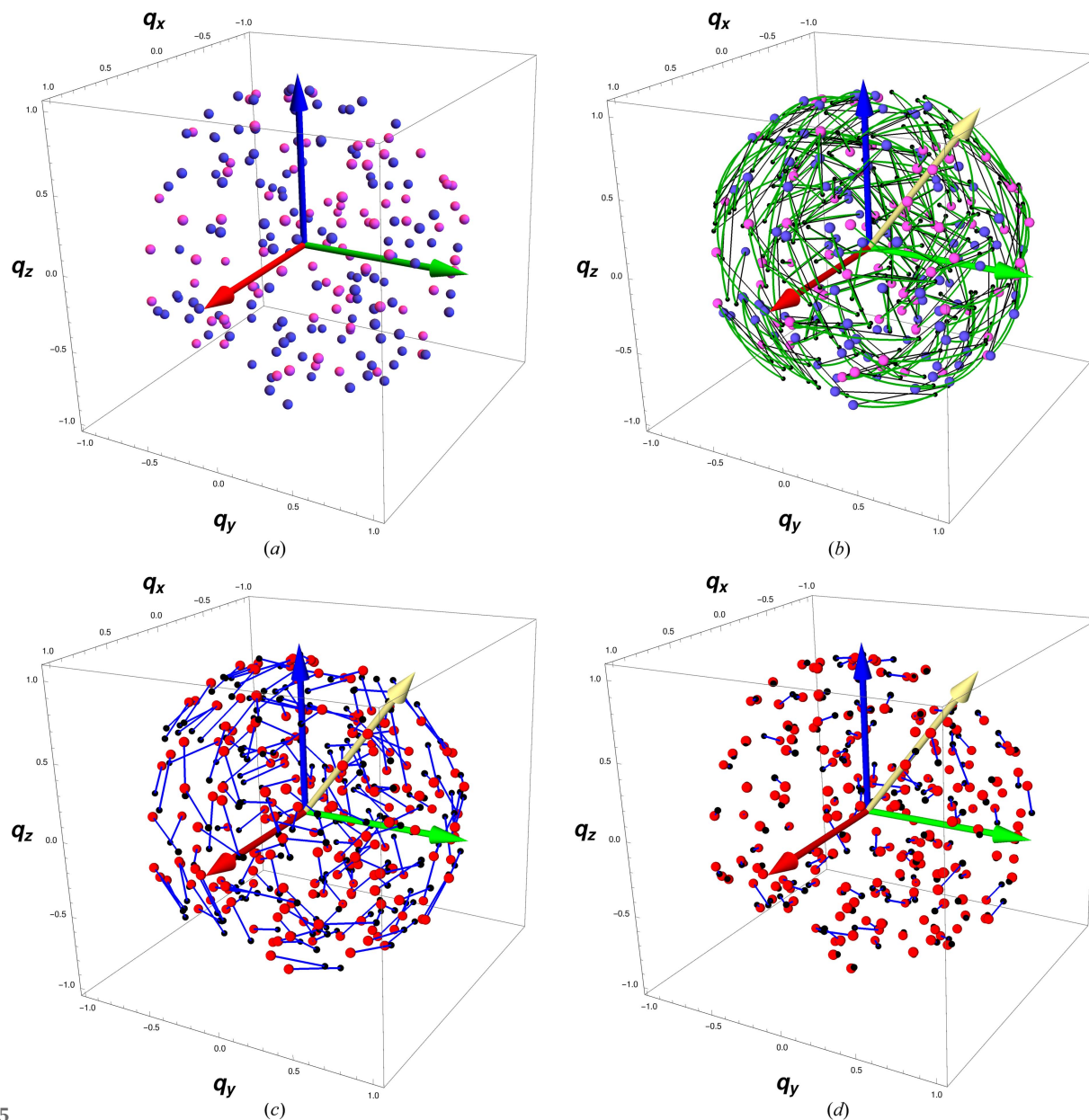


Figure 5

3D components of a quaternion orientation data set. (a) A quaternion reference set, color-coded by the sign of q_0 . (b) Exact quaternion arc-length distances (green arcs) versus chord distances (black lines) between the test points (black dots) and the reference points. (c) Part way from the starting state to the aligned state, at $s = 0.5$. (d) The final best alignment at $s = 1.0$. The yellow arrow is the direction of the quaternion eigenvector; when scaled, the length is the sine of half the optimal rotation angle.

quaternion, and adjust the sign of each data value t_k to \tilde{t}_k by the transformation

$$\tilde{t}_k = t_k \operatorname{sign}(q \cdot t_k) \rightarrow |q \cdot t_k| = q \cdot \tilde{t}_k. \quad (42)$$

The neighborhood of q matters because, as argued by Hartley *et al.* (2013), even though the allowed range of 3D rotation angles is $\theta \in (-\pi, \pi)$ [or quaternion sphere angles $\alpha \in (-\pi/2, \pi/2)$], convexity of the optimization problem cannot be guaranteed for collections outside local regions centered on some θ_0 of size $\theta_0 \in (-\pi/2, \pi/2)$ [or $\alpha_0 \in (-\pi/4, \pi/4)$]: beyond this range, local basins may exist that allow the mapping equation (42) to produce distinct local variations in the assignments of the $\{\tilde{t}_k\}$ and in the solutions for q_{opt} . Within

considerations of such constraints, equation (42) now allows us to take the summation outside the absolute value, and write the quaternion-frame optimization problem in terms of maximizing the cross-term expression

$$\left. \begin{aligned} \Delta_{\text{chord}}(q) &= \sum_{k=1}^N q \cdot \tilde{t}_k \\ &= q \cdot V(t) \end{aligned} \right\}, \quad (43)$$

where $V = \sum_{k=1}^N \tilde{t}_k$ is proportional to the mean of the quaternion displacements $\{\tilde{t}_k\}$, defining their chord-distance *quaternion average*. V also clearly plays a role analogous to the Euclidean RMSD profile matrix M . However, since equation

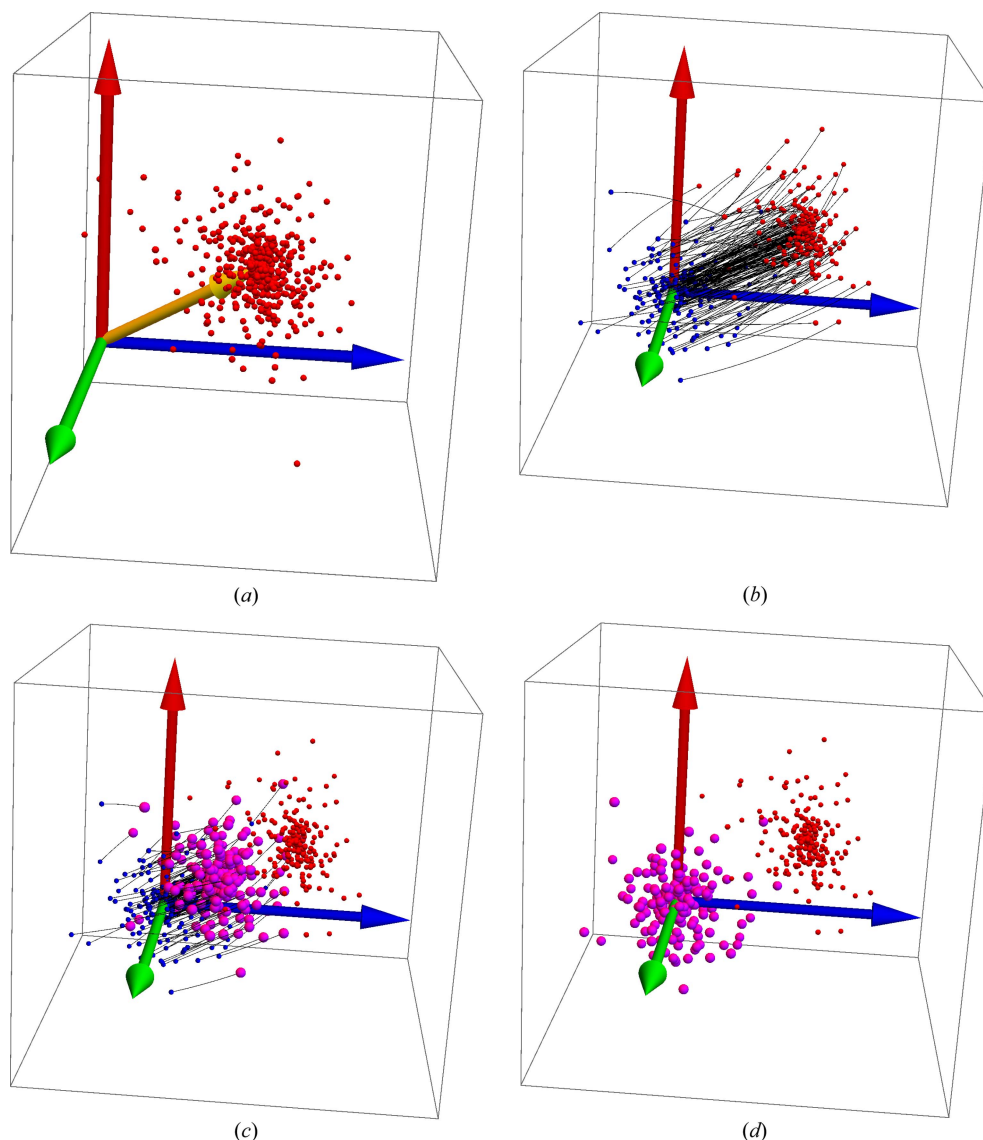


Figure 6
 3D components of the rotation-average transformation of the quaternion orientation data set, with each point denoting the *displacement* between each pair of frames as a single quaternion, corresponding to the rotation taking the test frame to the reference frame. (a) The cluster of points $t_k = r_k \star \tilde{p}_k \rightarrow \tilde{t}_k$ derived from the frame-matching problem using just the curved arcs in Fig. 5(b). If there were no alignment errors introduced in the simulation, these would all be a single point. The yellow arrow is the quaternion solution to the chord-distance centroid of this cluster and is identical to the optimal quaternion rotation transforming the test data to have the minimal chord measure relative to the reference data. (b) Choosing a less-cluttered subset of the data in (a), we display the geodesic paths from the initial quaternion displacements \tilde{t}_k to the origin-centered set with minimal chord-measure distance relative to the origin. This is the result of applying the inverse of the quaternion q_{opt} to each \tilde{t}_k . Note that the paths are curved geodesics lying properly within the quaternion sphere. (c, d) Rotating the cluster using a slerp between the quaternion barycenter of the initial misaligned data and the optimally aligned position, which is centered at the origin.

(43) is *linear* in q , we have the remarkable result that, as noted in the treatment of Hartley *et al.* (2013) regarding the quaternion L_2 chordal-distance norm, the solution is immediate, being simply

$$q_{opt} = \frac{V}{\|V\|}, \tag{44}$$

since that obviously maximizes the value of $\Delta_{chord}(q)$. This gives the maximal value of the measure as

$$\Delta_{chord}(q_{opt}) = \|V\| \tag{45}$$

and thus $\|V\|$ is the exact orientation-frame analog of the spatial RMSD maximal eigenvalue ϵ_{opt} , except it is far easier to compute.

7.4. Illustrative example

Using the quaternion display method described in Appendix B and illustrated in Fig. 12, we present in Fig. 5(a) a representative quaternion-frame reference data set, then in (b) we include a matching set of rotated noisy test data (small black dots), and draw the arc and chord distances (see also Fig. 4) connecting each test-reference point pair in the quaternion

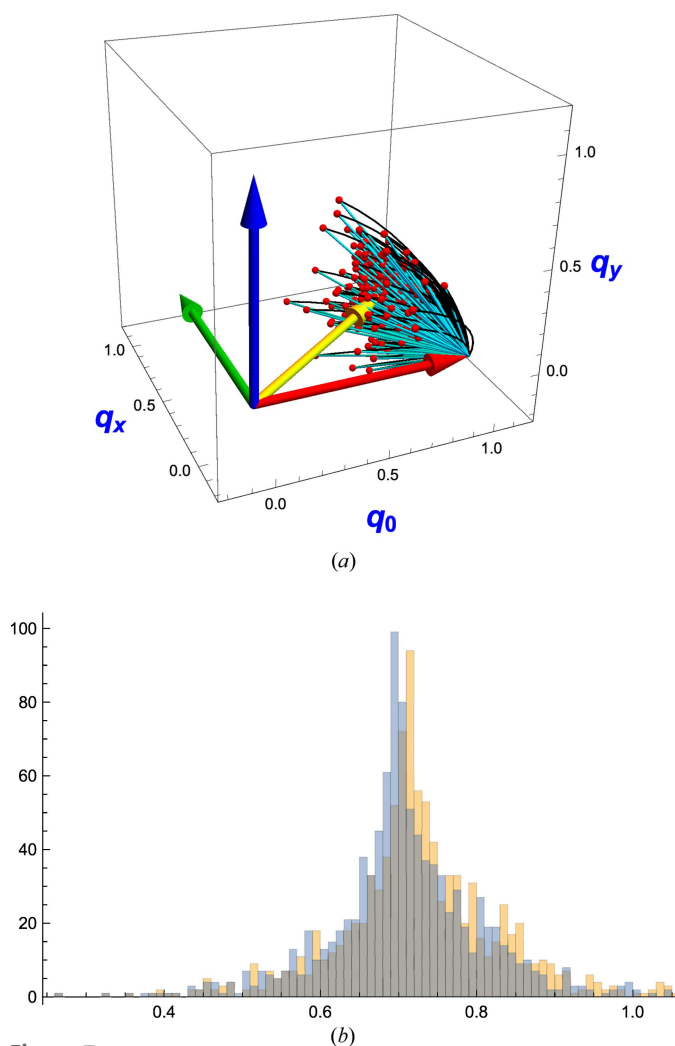


Figure 7

(a) Projecting the geodesic versus chord distances from the origin to sampled points in a set of frame-displacement data $t_k = r_k \star \bar{p}_k \rightarrow \tilde{t}_k$. Since the \mathbf{q} spatial quaternion paths project to a straight line from the origin, we use the (q_0, q_x, q_y) coordinates instead of our standard \mathbf{q} coordinates to expose the curvature in the arc-length distances to the origin. (b) Histogram of the chord-length distances to the origin (in blue) compared to the histogram of the geodesic arc-length distances (in yellow), sampled using a uniform distribution of random quaternions over a portion of \mathbf{S}^3 . If there were no errors, all the points would have the same distance from the origin located at $q = (1, 0, 0, 0)$ [red axis in (a)], and there would be one blue spike, appearing at a slightly smaller position than the yellow spike because arc length is always longer than chord length. The arc-length method has a different distribution, as expected, and produces a very slightly better barycenter. However, the optimal quaternions for the arc-length versus chord-length measure for this simulated data set differ by only a fraction of a degree, so drawing the positions of the two distinct optimal quaternions would not reveal any noticeable difference in image (a).

space. In Fig. 5(c,d), we show the results of the quaternion-frame alignment process using conceptually the same slerp of equation (15) to transition from the raw state at $q(s=0) = q_{\text{ID}}$ to $q(s=0.5)$ for (c) and $q(s=1.0) = q_{\text{opt}}$ for (d). The yellow arrow is the axis of rotation specified by the spatial part of the optimal quaternion.

The rotation-averaging visualization of the optimization process, though it has exactly the same optimal quaternion, is

quite different, since all the quaternion data collapse to a list of single small quaternions $t = r \star \bar{p}$. As illustrated in Fig. 6, with compatible sign choices, the \tilde{t}_k 's cluster around the optimal quaternion, which is clearly consistent with being the barycenter of the quaternion differences, intuitively the place to which all the quaternion frames need to be rotated to optimally coincide. As before, the yellow arrow is the axis of rotation specified by the spatial part of the optimal quaternion. Next, Fig. 7 addresses the question of how the rigorous arc-length measure is related to the chord-length measure that can be treated using the same methods as the spatial RMSD optimization. In parallel to Fig. 5(b), Fig. 7(a) shows essentially the same comparison for the \tilde{t}_k quaternion-displacement version of the same data. In Fig. 7(b), we show the histograms of the chord distances to a sample point, the origin in this case, versus the arc-length or geodesic distances. They obviously differ, but in fact for plausible simulations, the arc-length numerical optimal quaternion barycenter differs from the chord-length counterpart by only a fraction of a degree. These issues are studied in more detail in the supporting information.

Next, in Fig. 8, we display the values of $\Delta_{\text{chord}} = q \cdot V$ that parallel the RMSD version in Fig. 3. The dots show the size of the cost $\Delta(q)$ at randomly sampled points across the entire \mathbf{S}^3 , with $q_0 \geq 0$ in (a) and $q_0 \leq 0$ in (b). We have all the signs of the \tilde{t}_k chosen to be centered in an appropriate local neighborhood, and so, unlike the quadratic Euclidean RMSD case, there is only one value for q_{opt} , which is in the direction of V . Finally, in Fig. 9 we present an intuitive sketch of the convexity constraints for the QFA optimization related to Hartley *et al.* (2013). We start with a set of data in (a) [with both $(q, -q)$ partners] that consists of three local clouds that can be smoothly deformed from dispersed to coinciding locations. Fig. 9(b) and (c) both contain a uniform sample of quaternion sample points q spread over all of quaternion space, shown as magenta dots, with positive and negative q_0 plotted on top of each other. Then each sample q is used to compute *one* set of mappings $t_k \rightarrow \tilde{t}_k$ and the *one* value of $q_{\text{opt}} = V(\tilde{t})/\|V\|$ that results. The black arrows show the relation of q_{opt} to each original sample q , effectively showing us their *votes* for the best quaternion average. Fig. 9(b) has the clusters positioned far enough apart that we can clearly see that there are several basins of attraction, with no unique solution for q_{opt} , while in (c), we have interpolated the three clusters to lie in the same local neighborhood, roughly in a ball of quaternion radius $\alpha < \pi/4$, and we see that almost all of the black arrows vote for one unique q_{opt} or its equivalent negative. This seems to be a useful exercise to gain intuition about the nature of the basins of attraction for the quaternion-averaging problem that is essential for quaternion-frame alignment.

7.5. Alternative matrix forms of the linear vector chord distance

If the signs of the quaternions representing orientation frames are well behaved, and the frame problem is our only concern, equations (43) and (44) provide a simple solution to

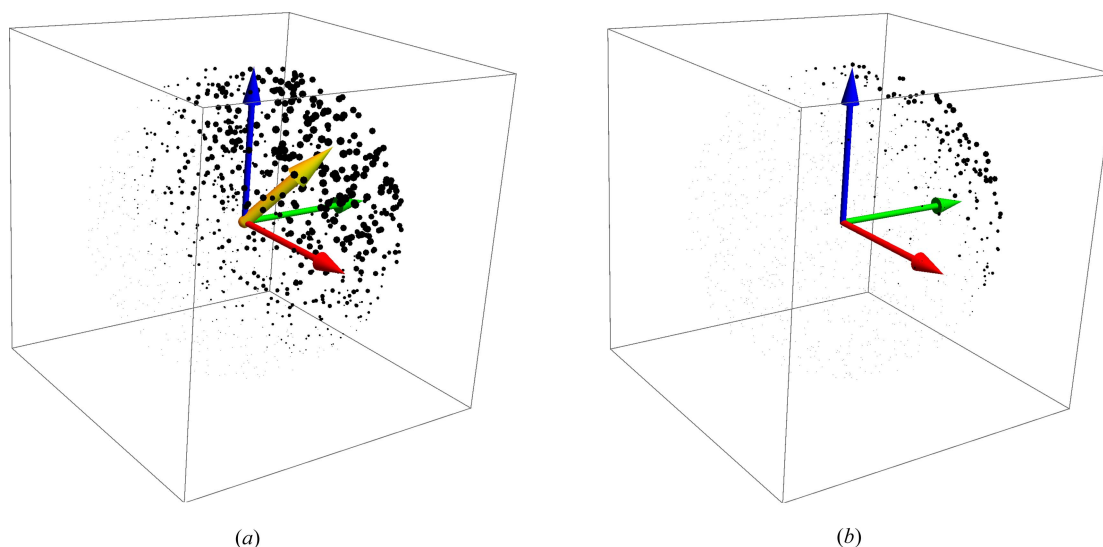


Figure 8 The values of $\Delta = q \cdot V$ represented by the sizes of the dots placed at a random distribution of quaternion points. We display the data dots at the locations of their spatial quaternion components \mathbf{q} . (a) is the northern hemisphere of \mathbf{S}^3 , with $q_0 \geq 0$, (b) is the southern hemisphere, with $q_0 \leq 0$, and we implicitly know that the value of q_0 is $\pm(1 - \mathbf{q} \cdot \mathbf{q})^{1/2}$. The points in these two solid balls represent the entire space of quaternions, and it is important to note that, even though $R(q) = R(-q)$ so each ball alone actually represents all possible unique rotation matrices, our cost function covers the *entire* space of quaternions, so q and $-q$ are distinct. The spatial component of the maximal eigenvector is shown by the yellow arrow, which clearly ends in the middle of the maximum values of Δ . The small cloud at the edge of (b) is simply the rest of the complete cloud around the tip of the yellow arrow, as q_0 passes through the ‘equator’ at $q_0 = 0$, going from a small positive value at the edge of (a) to a small negative value at the edge of (b).

finding the optimal global rotation. If we are anticipating wanting to combine a spatial profile matrix $M(E)$ with an orientation problem in a single 4×4 matrix, or we have problems defining a consistent quaternion sign, there are two further choices of orientation-frame measure we may consider.

(1) *Matrix form of the linear vector chord distance.* The first option uses the fact that the square of equation (43) will yield the same extremal solution for q_{opt} , so we can choose a measure of the form

$$\begin{aligned} \Delta_{\text{chord-sq}} &= (q \cdot V)(q \cdot V) \\ &= \sum_{a=0, b=0}^3 q_a V_a V_b q_b \\ &= q \cdot \Omega \cdot q, \end{aligned} \tag{46}$$

where $\Omega_{ab} = V_a V_b$ is a 4×4 rank-one symmetric matrix with $\det \Omega = 0$, and $\text{tr} \Omega = \sum_a V_a^2 \neq 0$. The eigensystem of Ω is just defined by the eigenvalue $\|V\|^2$, and combination with the spatial eigensystem can be achieved either numerically or algebraically. The sign issues for the sampled data remain unchanged since they appear inside the sums defining V . This form will acquire more importance in the 4D case.

(2) *Fixing the sign problem with the quadratic rotation matrix chord distance.* Our second approach has a very natural way to *eliminate sign dependence altogether* from the quaternion chord-distance method, and has a close relationship to Δ_{chord} . This measure is constructed starting from a minimized Fröbenius norm of the form [this approach is used by Sarlette & Sepulchre (2009); see also, *e.g.*, Huynh (2009), as well as

Moakher (2002), Markley *et al.* (2007), and Hartley *et al.* (2013)]

$$\|R(q) \cdot R(p_k) - R(r_k)\|_{\text{Frob}}^2$$

and then reducing to the cross-term as usual. The cross-term measure to be maximized, in terms of 3×3 (quaternion-sign-independent) rotation matrices, then becomes

$$\begin{aligned} \Delta_{\text{RRR}} &= \sum_{k=1}^N \text{tr}[R(q) \cdot R(p_k) \cdot R^{-1}(r_k)] = \sum_{k=1}^N \text{tr}[R(q \star p_k \star \bar{r}_k)] \\ &= \sum_{k=1}^N \text{tr}[R(q) \cdot R(p_k \star \bar{r}_k)] = \sum_{k=1}^N \text{tr}[R(q) \cdot R^{-1}(r_k \star \bar{p}_k)], \end{aligned} \tag{47}$$

where \bar{r} denotes the complex conjugate or inverse quaternion. We can verify that this is a chord distance by noting that each relevant $R \cdot R \cdot R$ term reduces to the square of an individual chord distance appearing in Δ_{chord} :

$$\left. \begin{aligned} &\sum_{k=1}^N \text{tr}[R(q) \cdot R(p_k) \cdot R(\bar{r}_k)] \\ &= \sum_{k=1}^N (4((q \star p_k) \cdot r_k)^2 - (q \cdot q)(p_k \cdot p_k)(r_k \cdot r_k)) \\ &= \sum_{k=1}^N (4(q \cdot (r_k \star \bar{p}_k))^2 - 1) \\ &= 4 \sum_{a,b} q_a \left(\sum_{k=1}^N [t_k]_a [t_k]_b \right) q_b - N \\ &= 4q \cdot A(t = r \star \bar{p}) \cdot q - N. \end{aligned} \right\} \tag{48}$$

Here the non-conjugated ordinary r on the right-hand side is not a typographical error, and the 4×4 matrix $A(t)$ is the alternative (equivalent) profile matrix that was introduced by

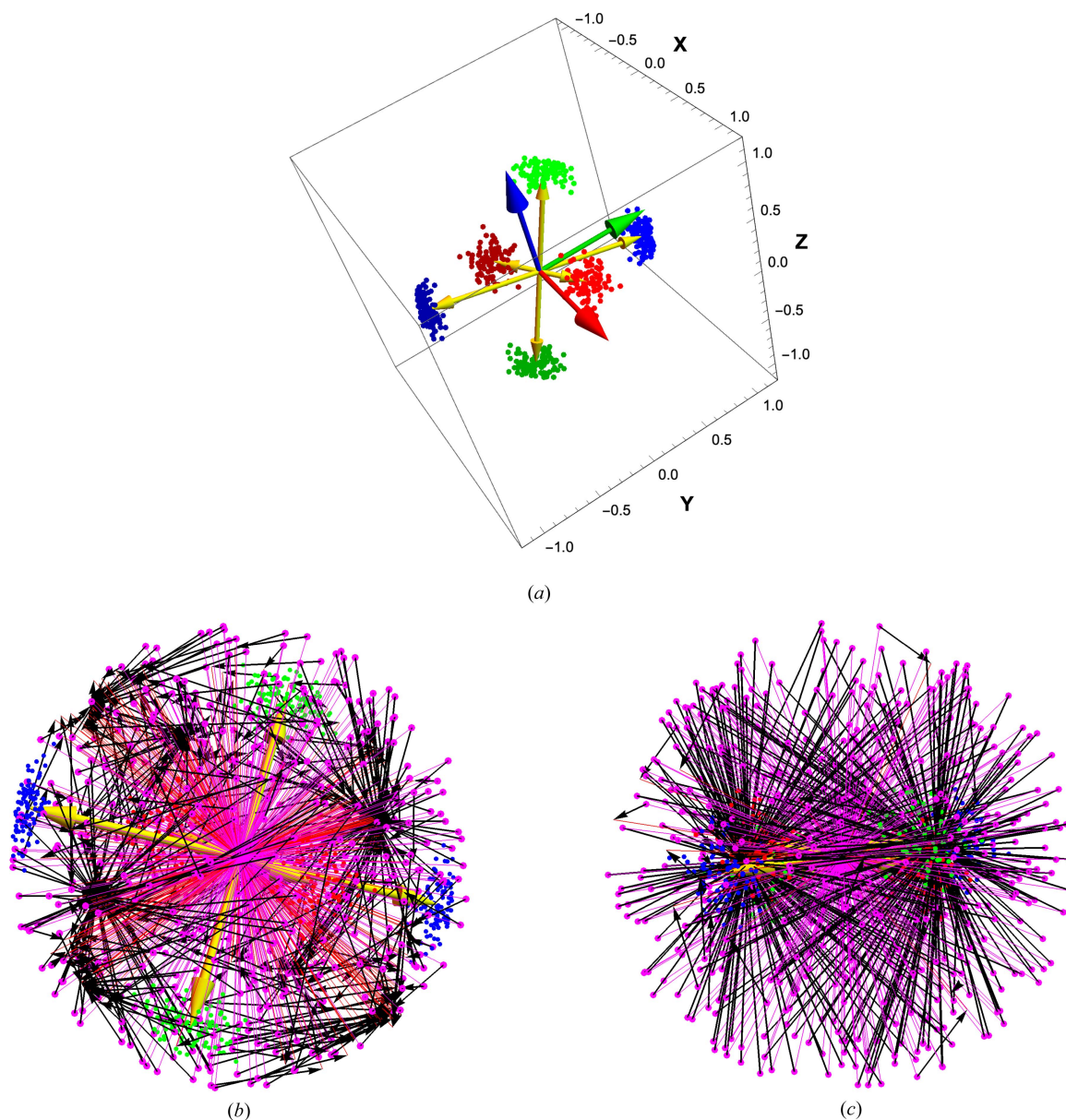


Figure 9

The behavior of the basins of attraction for the $t_k \rightarrow \tilde{t}_k$ map is shown here, starting in (a) with the $(q, -q)$ pairs for three movable clusters of quaternion-frame data, each having a well defined *local* quaternion average $q_{\text{opt}} = V(\tilde{t})/\|V\|$ shown as the yellow arrows with their $q \rightarrow -q$ equivalents. Next we merge all three samples into one data set that can be smoothly interpolated between the data being outside the $\alpha = \pi/4$ safe zone to all being together within that geometric boundary in quaternion space. Part (b) shows the results of taking 500 uniform samples of q and computing the set $\{\tilde{t}_k\}$ for *each* sample q , placed at the magenta dots, and *then* computing the resulting q_{opt} ; the black arrows follow the line from the sample point to the resultant q_{opt} . Clearly in (b), where the clusters are in their initial widely dispersed configuration, the black arrows (the ‘votes’ for the best q_{opt}) collect in several different basins of attraction, signifying the absence of a global solution. We then interpolate all the clusters close to each other, and show the new results of the voting in (c). Now almost all of the samplings of the full quaternion space converge to point their arrows densely to the two opposite values of q_{opt} , and there is just one effective basin of attraction.

Markley *et al.* (2007) and Hartley *et al.* (2013) for the chord-based quaternion-averaging problem. We can therefore use either the measure Δ_{RRR} or

$$\Delta_{\text{A}} = q \cdot A(t) \cdot q \quad (49)$$

with $A_{ab} = \sum_{k=1}^N [t_k]_a [t_k]_b$ as our rotation-matrix-based sign-insensitive chord-distance optimization measure. Exactly like our usual spatial measure, these measures must be *maximized* to find the optimal q . It is, however, important to emphasize that the optimal quaternion will *differ* for the Δ_{chord} , $\Delta_{\text{chord-sq}}$

and $\Delta_{\text{RRR}} \sim \Delta_{\text{A}}$ measures, though they will normally be very similar (see the discussion in the supporting information).

We now recognize that the sign-insensitive measures are all very closely related to our original spatial RMSD problem, and all can be solved by finding the optimal quaternion eigenvector q_{opt} of a 4×4 matrix. The procedure for $\Delta_{\text{chord-sq}}$ and Δ_{A} follows immediately, but it is useful to work out the options for Δ_{RRR} in a little more detail. Defining $T_k = R(p_k) \cdot R^{-1}(r_k) = R(p_k \star \bar{r}_k) = R^{-1}(t_k)$, we can write our optimization measure as

$$\begin{aligned}\Delta_{\text{RRR}} &= \sum_{k=1}^N \text{tr}(R(q) \cdot T_k) = \sum_{a=1, b=1}^3 R_{ba}(q) T_{ab} \\ &= \sum_{a=0, b=0}^3 q_a \cdot U_{ab}(p, r) \cdot q_b = q \cdot U(p, r) \cdot q, \quad (50)\end{aligned}$$

where the frame-based cross-covariance matrix is simply $T_{ab} = \sum_{k=1}^N [T_k]_{ab}$ and $U(p, r) = U(T)$ has the same relation to T as $M(E)$ has to E in equation (13).

To compute the necessary 4×4 numerical profile matrix U , one need only substitute the appropriate 3D frame triads or their corresponding quaternions for the k th frame pair and sum over k . Since the orientation-frame profile matrix $U(p, r)$ is symmetric and traceless just like the Euclidean profile matrix M , the same solution methods for the optimal quaternion rotation q_{opt} will work without alteration in this case, which is probably the preferable method for the general problem.

7.6. Evaluation

The details of evaluating the properties of our quaternion-frame alignment algorithms, including comparison of the chord approximation to the arc-length measure, are available in the supporting information. The top-level result is that, even for quite large rotational differences, the mean difference between the optimal quaternion using the numerical arc-length measure and the optimal quaternion using the chord approximation for any of the three methods is on the order of small fractions of a degree for the random data distributions that we examined.

8. The 3D combined point + frame alignment problem

Since we now have precise alignment procedures for both 3D spatial coordinates and 3D frame triad data (using the exact measure for the former and the approximate chord measure for the latter), we can consider the full 6 degree-of-freedom (6DOF) alignment problem for combined data from a single structure. As always, this problem can be solved either by numerical eigenvalue methods or in closed algebraic form using the eigensystem formulation of both alignment problems presented in the previous sections. While there are clearly appropriate domains of this type, *e.g.* any protein structure in the Protein Data Bank can be converted to a list of residue centers and their local frame triads (Hanson & Thakur, 2012), little is known at this time about the potential value of combined alignment. To establish the most complete possible picture, we now proceed to describe the details of our solution to the alignment problem for combined translational and rotational data, but we remark at the outset that the results of the combined system are not obviously very illuminating.

The most straightforward approach to the combined 6DOF measure is to equalize the scales of our spatial $M(E)$ profile matrix and our orientation-frame $U(S)$ profile matrix by imposing a unit-eigenvalue normalization, and then simply to perform a linear interpolation modified by a dimensional

constant σ to adjust the relative importance of the orientation-frame portion:

$$\Delta_{xf}(t, \sigma) = q \cdot \left[(1-t) \frac{M(E)}{\epsilon_x} + t\sigma \frac{U(S)}{\epsilon_f} \right] \cdot q. \quad (51)$$

Because of the dimensional incompatibility of Δ_x and Δ_f , we treat the ratio

$$\lambda^2 = \frac{t\sigma}{1-t}$$

as a dimensional weight such as that adopted by Fogolari *et al.* (2016) in their entropy calculations, or implicit in the weights α and β employed in the error function of Walker *et al.* (1991). If we take t to be dimensionless, then σ carries the dimensional scale information.

Given the composite profile matrix of equation (51), we can now extract our optimal rotation solution by computing the maximal eigenvalue as usual, either numerically or algebraically (though we may need the extension to the non-vanishing trace case examined in the supporting information for some choices of U). The result is a parameterized eigensystem

$$\left. \begin{array}{l} \epsilon_{\text{opt}}(t, \sigma) \\ q_{\text{opt}}(t, \sigma) \end{array} \right\} \quad (52)$$

yielding the optimal values $R(q_{\text{opt}}(t, \sigma))$, $\Delta_{xf} = \epsilon_{\text{opt}}(t, \sigma)$ based on the data $\{E, S\}$ no matter what we take as the values of the two variables (t, σ) .

8.1. A simplified composite measure

However, upon inspection of equation (51), one wonders what happens if we simply use the slerp defined in equation (8) to interpolate between the *separate* spatial and orientation-frame optimal quaternions. While the eigenvalues that correspond to the two scaled terms M/ϵ_x and U/ϵ_f in equation (51) are both unity, and thus differ from the eigenvalues of M and U , the individual normalized eigenvectors $q_{x:\text{opt}}$ and $q_{f:\text{opt}}$ are the same. Thus, if we are happy with simply using a hand-tuned fraction of the combination of the two corresponding rotations, we can just choose a composite rotation $R(q(t))$ specified by

$$q(t) = \text{slerp}(q_{x:\text{opt}}, q_{f:\text{opt}}, t) \quad (53)$$

to study the composite 6DOF alignment problem. In fact, as detailed in the supporting information, if we simply plug this $q(t)$ into equation (51) for any t (and $\sigma = 1$), we find *negligible differences* between the quaternions $q(t)$ and $q_{\text{opt}}(t, 1)$ as a function of t . We suggest in addition that any particular effect of $\sigma \neq 1$ could be achieved at some value of t in the interpolation. We thus conclude that, for all practical purposes, we might as well use equation (53) with the parameter t adjusted to achieve the objective of equation (51) to study composite translational and rotational alignment similarities.

9. Conclusion

Our objective has been to explore quaternion-based treatments of the RMSD data-comparison problem as developed in

the work of Davenport (1968), Faugeras & Hebert (1983), Horn (1987), Diamond (1988), Kearsley (1989), and Kneller (1991), among others, and to publicize the exact algebraic solutions, as well as extending the method to handle wider problems. We studied the intrinsic properties of the RMSD problem for comparing spatial-coordinate and orientation-frame data in quaternion-accessible domains, and we examined the nature of the solutions for the eigensystems of the 3D spatial-coordinate RMSD problem, as well as the corresponding 3D quaternion orientation-frame alignment problem (QFA). Extensions of both the spatial-coordinate and orientation-frame alignment problems and their solutions to 4D are detailed in the supporting information. We also examined solutions for the combined 3D spatial-coordinate and orientation-frame RMSD problem, arguing that a simple quaternion interpolation between the two individual solutions may well be sufficient for most purposes.

APPENDIX A The 3D Euclidean-space least-squares matching function

This appendix works out the details of the long-form least-squares distance measure for the 3D Euclidean alignment problem using the method of Hebert and Faugeras (Hebert, 1983; Faugeras & Hebert, 1983, 1986). Starting with the 3D Euclidean minimizing distance measure equation (9), we can exploit equation (5) for $R(q)$, along with equation (2), to produce an alternative quaternion eigenvalue problem whose *minimal* eigenvalue determines the eigenvector q_{opt} specifying the matrix that rotates the test data into closest correspondence with the reference data.

Adopting the convenient notation $\mathbf{x} = (0, x_1, x_2, x_3)$ for a pure imaginary quaternion, we employ the following steps:

$$\begin{aligned} S_3 &= \sum_{k=1}^N \|R_3(q) \cdot x_k - y_k\|^2 \\ &= \sum_{k=1}^N \|q \star \mathbf{x}_k \star \bar{q} - \mathbf{y}_k\|^2 = \sum_{k=1}^N \|q \star \mathbf{x}_k \star \bar{q} - \mathbf{y}_k\|^2 \|q\|^2 \\ &= \sum_{k=1}^N \|q \star \mathbf{x}_k - \mathbf{y}_k \star q\|^2 \quad \text{by equation (2)} \\ &= \sum_{k=1}^N \|A(\mathbf{x}_k, \mathbf{y}_k) \cdot q\|^2 = \sum_{k=1}^N q \cdot A_k^T \cdot A_k \cdot q \\ &= \sum_{k=1}^N q \cdot B_k \cdot q = q \cdot B \cdot q. \end{aligned} \quad (54)$$

Here we may write, for each k , the matrix $A(\mathbf{x}_k, \mathbf{y}_k)$ as

$$A_k = \begin{bmatrix} 0 & -a_1 & -a_2 & -a_3 \\ a_1 & 0 & s_3 & -s_2 \\ a_2 & -s_3 & 0 & s_1 \\ a_3 & s_2 & -s_1 & 0 \end{bmatrix}_k \quad (55)$$

where, with ‘ a ’ for ‘antisymmetric’ and ‘ s ’ for ‘symmetric,’

$$\begin{aligned} a_{\{1,2,3\}} &= \{x_1 - y_1, x_2 - y_2, x_3 - y_3\} \\ s_{\{1,2,3\}} &= \{x_1 + y_1, x_2 + y_2, x_3 + y_3\} \end{aligned}$$

and, again for each k ,

$$B_k = A_k^T \cdot A_k = \begin{bmatrix} a_1^2 + a_2^2 + a_3^2 & a_3s_2 - a_2s_3 & a_1s_3 - a_3s_1 & a_2s_1 - a_1s_2 \\ a_3s_2 - a_2s_3 & a_1^2 + s_2^2 + s_3^2 & a_1a_2 - s_1s_2 & a_1a_3 - s_1s_3 \\ a_1s_3 - a_3s_1 & a_1a_2 - s_1s_2 & a_2^2 + s_1^2 + s_3^2 & a_2a_3 - s_2s_3 \\ a_2s_1 - a_1s_2 & a_1a_3 - s_1s_3 & a_2a_3 - s_2s_3 & a_3^2 + s_1^2 + s_2^2 \end{bmatrix}_k \quad (56)$$

and $B = \sum_{k=1}^N B_k$. As using the full squared-difference minimization measure equation (9) requires the global minimal value, the solution for the optimal quaternion in equation (54) is the eigenvector of the *minimal* eigenvalue of B in equation (56). This is the approach used by Faugeras and Hebert in the earliest application of the quaternion method to scene alignment that we know of. While it is important to be aware of this alternative method, in the main text we have found it more useful to focus on the alternate form exploiting only the non-constant cross-term appearing in equation (9), as does most of the recent molecular structure literature. The cross-term requires the determination of the *maximal* eigenvalue rather than the *minimal* eigenvalue of the corresponding data matrix. Direct numerical calculation verifies that, though the minimal eigenvalue of equation (56) differs from the maximal eigenvalue of the cross-term approach, the exact same optimal eigenvector is obtained, a result that can presumably be proven algebraically but that we will not need to pursue here.

APPENDIX B Introduction to quaternion orientation frames

B1. What is a quaternion frame?

We will first present a bit of intuition about coordinate frames that may help some readers with our terminology. If we take the special case of a quaternion representing a rotation in the 2D (x, y) plane, the 3D rotation matrix equation (5) reduces to the standard right-handed 2D rotation

$$R_2(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}. \quad (57)$$

As shown in Fig. 10(a), we can use θ to define a unit direction in the complex plane defined by $z = \exp i\theta$, and then the *columns* of the matrix $R_2(\theta)$ naturally correspond to a unique associated 2D coordinate frame diad; an entire collection of points z and their corresponding frame diads are depicted in Fig. 10(b).

Starting from this context, we can get a clear intuitive picture of what we mean by a ‘quaternion frame’ before diving into the quaternion RMSD problem. The essential step is to look again at equation (5) for $n_x = 1$, and write the corresponding quaternion as $(a, b, 0, 0)$ with $a^2 + b^2 = 1$, so this is a ‘2D quaternion’, and is indistinguishable from a complex phase like $z = \exp i\theta$ that we just introduced. There is one

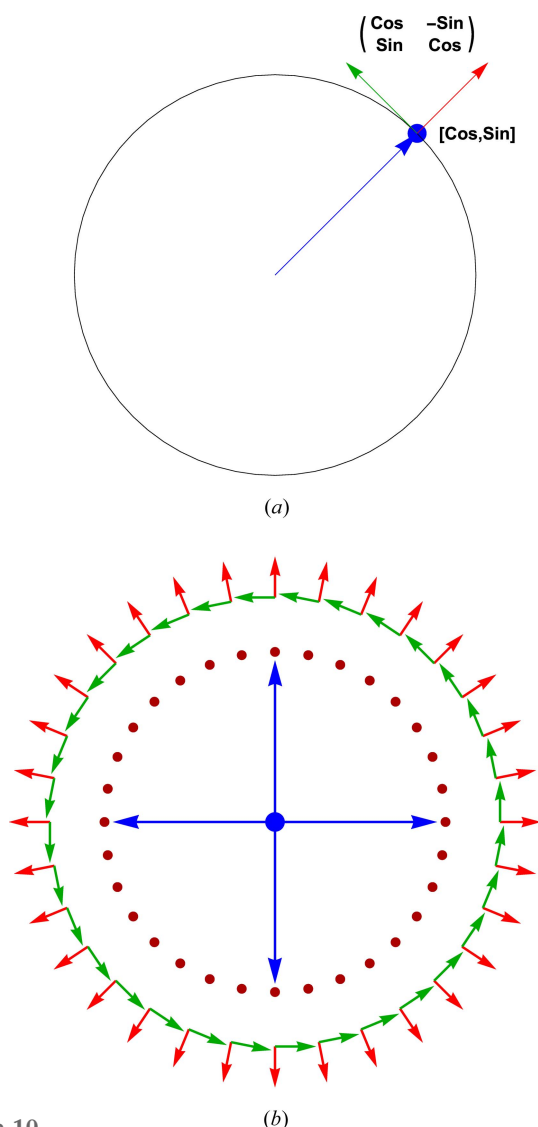


Figure 10
 (a) Any standard 2D coordinate frame corresponds to the columns of an ordinary rotation matrix, and is associated to the point $(\cos \theta, \sin \theta)$ on a unit circle. (b) The standard 2D coordinate frames associated with a sampling of the entire circle of points $(\cos \theta, \sin \theta)$.

significant difference, however, and that is that equation (5) shows us that $R_2(\theta)$ takes a new form, quadratic in a and b ,

$$R_2(a, b) = \begin{bmatrix} a^2 - b^2 & -2ab \\ 2ab & a^2 - b^2 \end{bmatrix}. \quad (58)$$

Using either the formula (7) for $q(\theta, \hat{\mathbf{n}})$ or just exploiting the trigonometric double-angle formulas, we see that equation (57) and equation (58) correspond and that

$$(a, b) = (\cos(\theta/2), \sin(\theta/2)) \quad (59)$$

$$u = (a + ib) = \sqrt{z} = \exp(i\theta/2). \quad (60)$$

Our simplified 2D quaternion thus describes the *square root* of the usual Euclidean frame given by the columns of $R_2(\theta)$. Thus the pair (a, b) (the reduced quaternion) itself corresponds to a frame. In Fig. 11(a), we show how a given ‘quaternion frame’, *i.e.* the columns of $R_2(a, b)$, corresponds to a point $u = a + ib$ in

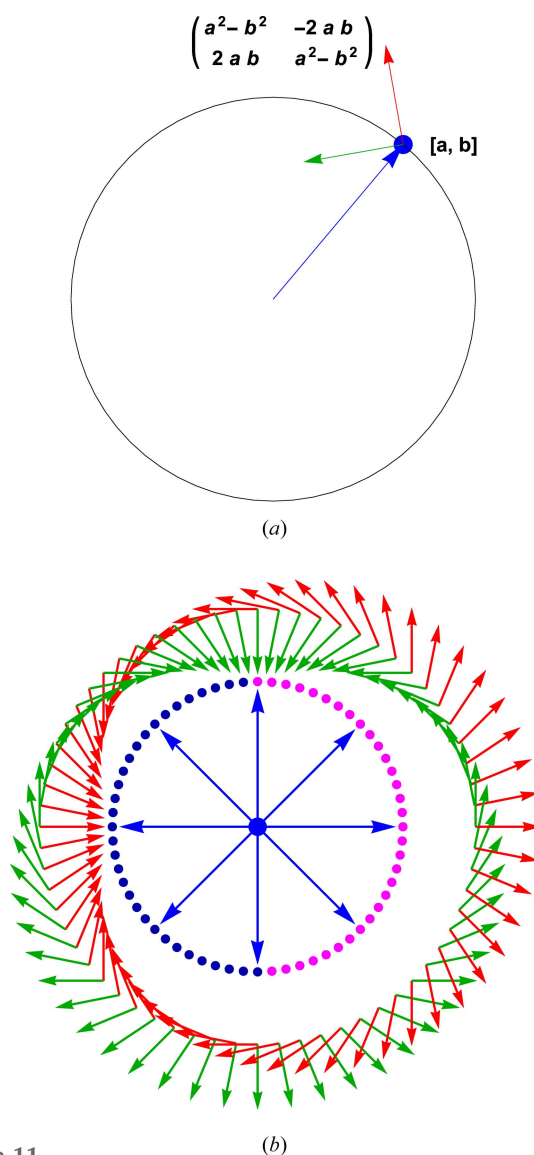


Figure 11
 (a) The quaternion point (a, b) , in contrast, corresponds via the double-angle formula to coordinate frames that rotate twice as rapidly as (a, b) progresses around the unit circle that is a simplified version of quaternion space. (b) The set of 2D frames associated with the entire circle of quaternion points (a, b) ; each diametrically opposite point corresponds to an identical frame. For later use in displaying full quaternions, we show how color coding can be used to encode the sign of one of the coordinates on the circle.

the complex plane. Diametrically *opposite* points (a, b) and $(-a, -b)$ now correspond to the *same* frame! Fig. 11(b) shows the corresponding frames for a large collection of points (a, b) in the complex plane, and we see the new and unfamiliar feature that the frames make *two* full rotations on the complex circle instead of just one as in Fig. 10(b).

This is what we have to keep in mind as we now pass to using a full quaternion to represent an arbitrary 3D frame triad via equation (5). The last step is to notice that in Fig. 11(b) we can represent the set of frames in one half of the complex circle, $a \geq 0$ shown in magenta, as distinct from those in the other half, $a < 0$ shown in dark blue; for any value of b , the vertical axis, there is a *pair* of a 's with opposite signs and

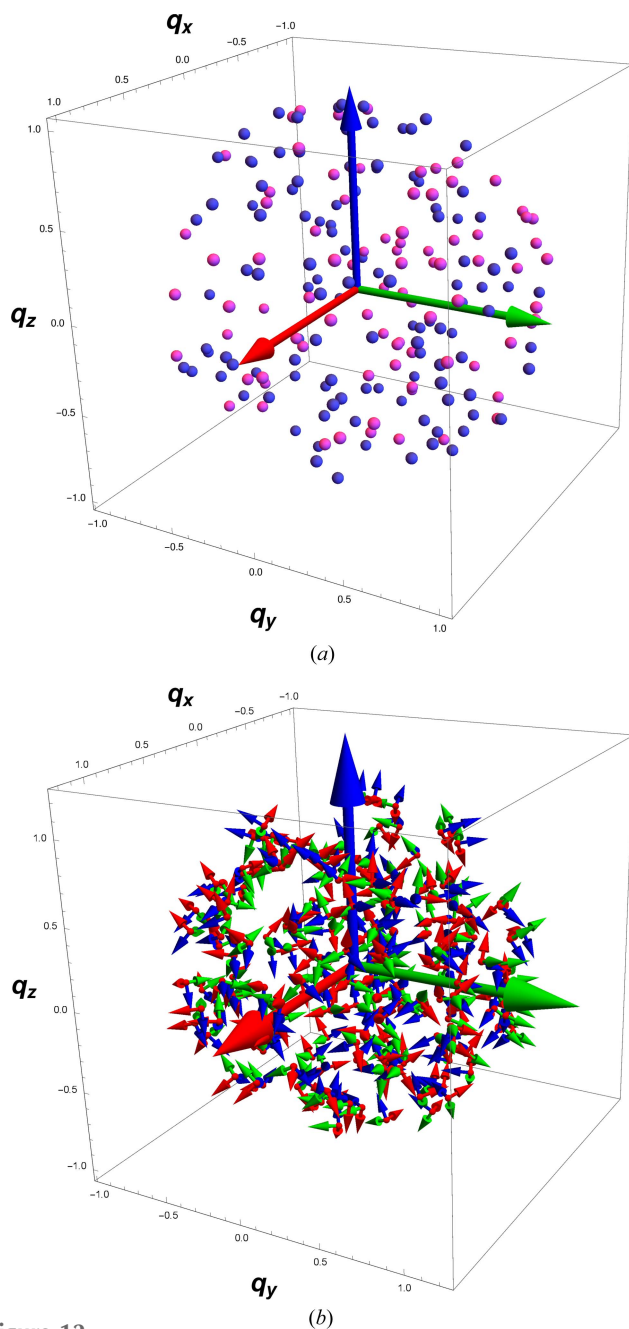


Figure 12

(a) The 3D portions of the quaternion reference-frame data $q = (q_0, q_x, q_y, q_z)$, using different colors for $q_0 \geq 0$ and $q_0 < 0$ in the unseen direction. Since $|q_0| = (1 - q_x^2 - q_y^2 - q_z^2)^{1/2}$, the complete quaternion can in principle be determined from the 3D display. (b) The 3D orientation-frame triads for each reference point (q_0, q_x, q_y, q_z) displayed at their associated $\mathbf{q} = (q_x, q_y, q_z)$.

colors. In the quaternion case, we can display quaternion frames inside one single sphere, like displaying only the b coordinates in Fig. 11(b) projected to the vertical axis, realizing that if we know the sign-correlated coloring, we can determine both the magnitude of the dependent variable $a = \pm(1 - b^2)^{1/2}$ as well as its sign. The same holds true in the general case: if we display only a quaternion's 3-vector part $\mathbf{q} = (q_x, q_y, q_z)$ along with a color specifying the sign of q_0 , we implicitly know both the magnitude and sign of

$q_0 = \pm(1 - q_x^2 - q_y^2 - q_z^2)^{1/2}$, and such a 3D plot therefore accurately depicts *any* quaternion. Another alternative employed in the main text is to use *two* solid balls, one a 'northern hemisphere' for the $q_0 \geq 0$ components and the other a 'southern hemisphere' for the $q_0 \leq 0$ components. Each may be useful in different contexts.

B2. Example

We illustrate all this in Fig. 12(a), which shows a typical collection of quaternion reference-frame data displaying only the \mathbf{q} components of (q_0, \mathbf{q}) ; the $q_0 \geq 0$ data are mixed with the $q_0 < 0$ data, but are distinguished by their color coding. In Fig. 12(b), we show the frame triads resulting from applying equation (5) to each quaternion point and plotting the result at the associated point \mathbf{q} in the display.

APPENDIX C

On obtaining quaternions from rotation matrices

The quaternion RMSD profile matrix method can be used to implement a singularity-free algorithm to obtain the (sign-ambiguous) quaternions corresponding to numerical 3D and 4D rotation matrices. There are many existing approaches to the 3D problem in the literature [see, *e.g.*, Shepperd (1978), Shuster & Natanson (1993), or Section 16.1 of Hanson (2006)]. In contrast to these approaches, Bar-Itzhack (2000) has observed, in essence, that if we simply replace the data matrix E_{ab} by a numerical 3D orthogonal rotation matrix R , the numeric quaternion q that corresponds to $R_{\text{numeric}} = R(q)$, as defined by equation (5), can be found by solving our familiar maximal quaternion eigenvalue problem. The initially unknown optimal matrix (technically its quaternion) computed by maximizing the similarity measure is equivalent to a single-element quaternion barycenter problem, and the construction is designed to yield a best approximation to R itself in quaternion form. To see this, take $S(r)$ to be the sought-for optimal rotation matrix, with its own quaternion r , that must maximize the Bar-Itzhack measure. We start with the Fröbenius measure describing the match of two rotation matrices corresponding to the quaternion r for the unknown quaternion and the numeric matrix R containing the known 3×3 rotation matrix data:

$$\begin{aligned} S_{\text{BI}} &= \|S(r) - R\|_{\text{Frob}}^2 = \text{tr}([S(r) - R] \cdot [S^{\text{T}}(r) - R^{\text{T}}]) \\ &= \text{tr}(I_3 + I_3 - 2(S(r) \cdot R^{\text{T}})) \\ &= \text{const} - 2 \text{tr} S(r) \cdot R^{\text{T}}. \end{aligned}$$

Pulling out the cross-term as usual and converting to a maximization problem over the unknown quaternion r , we arrive at

$$\Delta_{\text{BI}} = \text{tr} S(r) \cdot R^{\text{T}} = r \cdot K(R) \cdot r, \quad (61)$$

where R is (approximately) an orthogonal matrix of numerical data and $K(R)$ is analogous to the profile matrix $M(E)$. Now S is an abstract rotation matrix and R is supposed to be a good numerical approximation to a rotation matrix, and thus the product $T = S \cdot R^{\text{T}}$ should also be a good approximation to an

SO(3) rotation matrix; hence that product itself corresponds closely to some axis $\hat{\mathbf{n}}$ and angle θ , where (supposing we knew R 's exact quaternion q)

$$\text{tr } S(r) \cdot R^\top(q) = \text{tr } T(r \star \bar{q}) = \text{tr } T(\theta, \hat{\mathbf{n}}) = 1 + 2 \cos \theta.$$

The maximum is obviously close to T being the identity matrix, with the ideal value at $\theta = 0$, corresponding to $S \approx R$. Thus if we find the maximal quaternion eigenvalue ϵ_{opt} of the profile matrix $K(R)$ in equation (61), our closest solution is well represented by the corresponding normalized eigenvector r_{opt} ,

$$q = r_{\text{opt}}. \quad (62)$$

This numerical solution for q will correspond to the targeted numerical rotation matrix, solving the problem. To complete the details of the computation, we replace the elements E_{ab} in equation (13) by a general orthonormal rotation matrix with columns $\mathbf{X} = (x_1, x_2, x_3)$, $\mathbf{Y} = (y_1, y_2, y_3)$ and $\mathbf{Z} = (z_1, z_2, z_3)$, scaling by $1/3$, thus obtaining the special 4×4 profile matrix K whose elements in terms of a known numerical matrix $R = [\mathbf{X}|\mathbf{Y}|\mathbf{Z}]$ (transposed in the algebraic expression for K owing to the R^\top) are

$$K(R) = \frac{1}{3} \begin{bmatrix} x_1 + y_2 + z_3 & y_3 - z_2 & z_1 - x_3 & x_2 - y_1 \\ y_3 - z_2 & x_1 - y_2 - z_3 & x_2 + y_1 & x_3 + z_1 \\ z_1 - x_3 & x_2 + y_1 & -x_1 + y_2 - z_3 & y_3 + z_2 \\ x_2 - y_1 & x_3 + z_1 & y_3 + z_2 & -x_1 - y_2 + z_3 \end{bmatrix}. \quad (63)$$

Determining the algebraic eigensystem of equation (63) is a non-trivial task. However, as we know, any orthogonal 3D rotation matrix $R(q)$, or equivalently, $R^\top(q) = R(\bar{q})$, can also be ideally expressed in terms of quaternions via equation (5), and this yields an alternate useful algebraic form:

$$K(q) = \frac{1}{3} \begin{bmatrix} 3q_0^2 - q_1^2 - q_2^2 - q_3^2 & 4q_0q_1 & & & & \\ 4q_0q_1 & -q_0^2 + 3q_1^2 - q_2^2 - q_3^2 & & & & \\ 4q_0q_2 & 4q_1q_2 & & & & \\ 4q_0q_3 & 4q_1q_3 & & & & \\ & 4q_0q_2 & 4q_0q_3 & & & \\ & 4q_1q_2 & 4q_1q_3 & & & \\ & -q_0^2 - q_1^2 + 3q_2^2 - q_3^2 & 4q_2q_3 & & & \\ & 4q_2q_3 & -q_0^2 - q_1^2 - q_2^2 + 3q_3^2 & & & \end{bmatrix}. \quad (64)$$

This equation then allows us to quickly prove that K has the correct properties to solve for the appropriate quaternion corresponding to R . First we note that the coefficients p_n of the eigensystem are simply constants,

$$p_1 = 0, \quad p_2 = -\frac{2}{3}, \quad p_3 = -\frac{8}{27}, \quad p_4 = -\frac{1}{27}.$$

Computing the eigenvalues and eigenvectors using the symbolic quaternion form, we see that the eigenvalues are constant, with maximal eigenvalue exactly one, and the eigenvectors are almost trivial, with the maximal eigenvector

being the quaternion q that corresponds to the (numerical) rotation matrix:

$$\epsilon = \{1, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}\} \quad (65)$$

$$r = \left\{ \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix}, \begin{bmatrix} -q_1 \\ q_0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -q_2 \\ 0 \\ q_0 \\ 0 \end{bmatrix}, \begin{bmatrix} -q_3 \\ 0 \\ 0 \\ q_0 \end{bmatrix} \right\}. \quad (66)$$

The first column is the quaternion r_{opt} , with $\Delta_{\text{BI}}(r_{\text{opt}}) = 1$. (This would be 3 if we had not divided by 3 in the definition of K .)

Alternate version. From the quaternion-barycenter work of Markley *et al.* (2007) and the natural form of the quaternion-extraction problem in 4D in the supporting information, we know that equation (64) actually has a much simpler form with the same unit eigenvalue and natural quaternion eigenvector. If we simply take equation (64) multiplied by 3, add the constant term $I_4 = (q_0^2 + q_1^2 + q_2^2 + q_3^2)I_4$, and divide by 4, we get a more compact quaternion form of the matrix, namely

$$K'(q) = \begin{bmatrix} q_0^2 & q_0q_1 & q_0q_2 & q_0q_3 \\ q_0q_1 & q_1^2 & q_1q_2 & q_1q_3 \\ q_0q_2 & q_1q_2 & q_2^2 & q_2q_3 \\ q_0q_3 & q_1q_3 & q_2q_3 & q_3^2 \end{bmatrix}. \quad (67)$$

This has vanishing determinant and trace $\text{tr } K' = 1 = -p_1$, with all other p_k coefficients vanishing, leading to an eigensystem identical to equation (64):

$$\epsilon = \{1, 0, 0, 0\} \quad (68)$$

$$r = \left\{ \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix}, \begin{bmatrix} -q_1 \\ q_0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -q_2 \\ 0 \\ q_0 \\ 0 \end{bmatrix}, \begin{bmatrix} -q_3 \\ 0 \\ 0 \\ q_0 \end{bmatrix} \right\}. \quad (69)$$

As elegant as this is, in practice our numerical input data are from the 3×3 matrix R itself, and not the quaternions, so we will almost always just use those numbers in equation (63) to solve the problem.

C1. Completing the solution

In typical applications, *the solution is immediate, requiring only trivial algebra.* The maximal eigenvalue is always known in advance to be unity for any valid rotation matrix, so we need only to compute the eigenvector from the numerical matrix equation (63) with unit eigenvalue. We simply compute any column of the adjugate matrix of $[K(R) - I_4]$, or solve the equivalent linear equations of the form

$$(K(R) - 1 * I_4) \cdot \begin{bmatrix} 1 \\ v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0, \quad q = r_{\text{opt}} = \text{normalize} \begin{bmatrix} 1 \\ v_1 \\ v_2 \\ v_3 \end{bmatrix}. \quad (70)$$

As always, one may need to check for degenerate special cases.

C2. Non-ideal cases

It is important to note, as emphasized by Bar-Itzhack, that if there are *significant errors* in the numerical matrix R , then the actual non-unit maximal eigenvalue of $K(R)$ can be computed numerically or algebraically as usual, and then that eigenvalue's eigenvector determines the *closest* normalized quaternion to the errorful rotation matrix, which can be very useful since such a quaternion always produces a valid rotation matrix.

In any case, *up to an overall sign*, r_{opt} is the desired numerical quaternion q corresponding to the target numerical rotation matrix $R = R(q)$. In some circumstances one is looking for a uniform statistical distribution of quaternions, in which case the overall sign of q should be chosen *randomly*.

The Bar-Itzhack approach solves the problem of extracting the quaternion of an arbitrary numerical 3D rotation matrix in a fashion that involves no singularities and only trivial testing for special cases, thus essentially making the traditional methods obsolete. The extension of Bar-Itzhack's method to the case of 4D rotations is provided in the supporting information.

APPENDIX D

On defining the quaternion barycenter

The notion of a *Riemannian barycenter* is generally associated with the work of Grove *et al.* (1974), and may also be referred to as the *Karcher mean* (Karcher, 1977), defined as the point that minimizes the sum of squared geodesic distances from the elements of a collection of fixed points on a manifold. The general class of such optimization problems has also been studied, *e.g.*, by Manton (2004). We are interested here in the case of quaternions, which we know are points on the spherical 3-manifold \mathbf{S}^3 defined by the unit-quaternion subspace of \mathbb{R}^4 restricted to $q \cdot q = 1$ for any point q in \mathbb{R}^4 . This subject has been investigated by a number of authors, with Brown & Worsey (1992) discussing the problems with this computation in 1992, and Buss & Fillmore (2001) proposing a solution applicable to computer-graphics 3D orientation interpolation problems in 2001, inspired to some extent by Shoemake's 1985 introduction of the quaternion *slerp* as a way to perform geodesic orientation interpolations in 3D using equation (5) for $R(q)$. There are a variety of methods and studies related to the quaternion-barycenter problem. In 2002, Moakher published a rigorous account on averaging in the group of rotations (Moakher, 2002), while subsequent treatments included the work by Markley *et al.* (2007), focusing on aerospace and astronomy applications, and the comprehensive review by Hartley *et al.* (2013), aimed in particular at the machine vision and robotics community, with additional attention to conjugate rotation averaging (the 'hand-eye calibration' problem in robotics) and multiple rotation averaging. While we have focused on measures starting from sums of squares that lead to closed-form optimization problems, Hartley *et al.* (2011) have carefully studied the utility of the

corresponding L_1 norm and the iterative Weiszfeld algorithm for finding its optimal solution numerically.

The task at hand is basically to extend the Bar-Itzhack algorithm to an entire collection of frames instead of a single rotation. We need to find an optimal rotation matrix $R(q)$ that corresponds to the quaternion point closest to the geodesic center of an *unordered* set of reference data. We already know that the case of the 'barycenter' of a single orientation frame is solved by the Bar-Itzhack algorithm of Appendix C, which finds the quaternion closest to a single item of rotation matrix data (the quaternion barycenter of a single rotation is itself). For two items of data, $R_1 = R(q_1)$ and $R_2 = R(q_2)$, the quaternion barycenter is determined by the *slerp* interpolator to be

$$q(q_1, q_2)_{\text{barycenter}} = q_1 \star (\bar{q}_1 \star q_2)^{1/2} = \text{slerp}(q_1, q_2, \frac{1}{2}).$$

For three or more items, no closed form is currently known.

We start with a data set of N rotation matrices $R(p_k)$ that are represented by the quaternions p_k , and we want $R(q)$ to be as close as possible to the set of $R(p_k)$. That rotation matrix, or its associated quaternion point, are the orientation-frame analogs of the Euclidean barycenter for a set of Euclidean points. As before, it is clear that the mathematically most justifiable measure employs the *geodesic arc length* on the quaternion sphere; but to the best of anyone's knowledge, there is no way to apply linear algebra to find the corresponding $R(q_{\text{opt}})$. Achieving a numerical solution to that problem is the task solved by Buss & Fillmore (2001), as well as a number of others, including, *e.g.*, Moakher (2002), Markley *et al.* (2007), and Hartley *et al.* (2013). The problem that we can understand algebraically is, once again, the approximate *chord measure*. For the case treated in Section 7, with the assumption that we can consistently use the quaternions themselves, we could find a solution using just the Euclidean chord average V/N from $V = \sum_{k=1}^N p_k$, so the optimal quaternion for the measure $\Delta_{\text{chord}} = q \cdot V$ was simply $q_{\text{opt}} = V/\|V\|$.

However, if that is not an option, and we require an average rotation that is based more directly on the sign-insensitive rotation matrices themselves, we need to use the Δ_{RRR} method of Section 7.5. This turns out to be essentially an extension of the Bar-Itzhack algorithm from a single rotation to a collection of rotations, and that is the approach we present here. Our starting point is the sign-insensitive Fröbenius measure $\|M\|^2 = \text{tr}(M \cdot M^T)$, giving us the following starting point:

$$\left. \begin{aligned} \mathbf{S}_{\text{barycenter}}(q) &= \sum_{k=1}^N \|R(q) - R(p_k)\|^2 \\ &= \sum_{k=1}^N \text{tr}([R(q) - R(p_k)] \cdot [R^T(q) - R^T(p_k)]) \\ &= \sum_{k=1}^N \text{tr}(2I_3 - 2R(q) \cdot R^T(p_k)) \\ &= \sum_{k=1}^N (6 - 2 \text{tr} R(q) \cdot R(\bar{p}_k)) \end{aligned} \right\} \quad (71)$$

Dropping the constants and converting as usual to maximize over the cross-term instead of minimizing the distance measure, we define a tentative spherical barycenter as the

- Diamond, R. (1988). *Acta Cryst.* **A44**, 211–216.
- Euler, L. (1733). *Commentarii academiae scientiarum imperialis Petropolitanae*, **6**, 216–231. <http://www.eulerarchive.org/pages/E030.html>.
- Faugeras, O. & Hebert, M. (1983). *Proc. 8th Joint Conf. Artificial Intell.* **2**, 996–1002. Morgan Kaufmann. <http://dl.acm.org/citation.cfm?id=1623516.1623603>.
- Faugeras, O. & Hebert, M. (1986). *Int. J. Robot. Res.* **5**, 27–52.
- Flower, D. (1999). *J. Mol. Graph. Model.* **17**, 238–244.
- Fogolari, F., Dongmo Fomthum, C. J., Fortuna, S., Soler, M. A., Corazza, A. & Esposito, G. (2016). *J. Chem. Theory Comput.* **12**, 1–8.
- Gibson, W. (1960). *Educ. Psychol. Meas.* **20**, 713–721.
- Golub, G. & van Loan, C. (1983). *Matrix Computations*, 1st ed., Section 12.4. Baltimore: Johns Hopkins University Press.
- Green, B. F. (1952). *Psychometrika*, **17**, 429–440.
- Grove, K., Karcher, H. & Ruh, E. A. (1974). *Math. Ann.* **211**, 7–21.
- Hanson, A. J. (2006). *Visualizing Quaternions*. Morgan-Kaufmann/Elsevier.
- Hanson, A. J. & Thakur, S. (2012). *J. Mol. Graph. Model.* **38**, 256–278.
- Hartley, R., Aftab, K. & Trampf, J. (2011). *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* pp. 3041–3048. IEEE Computer Society.
- Hartley, R., Trampf, J., Dai, Y. & Li, H. (2013). *Int. J. Comput. Vis.* **103**, 267–305.
- Havel, T. & Najfeld, I. (1994). *J. Mol. Struct. Theochem*, **308**, 241–262.
- Hebert, M. (1983). PhD thesis, University of Paris South. Available as INRIA Tech. Rep. ISBN 2-7261-0379-0.
- Horn, B. K., Hilden, H. M. & Negahdaripour, S. (1988). *J. Opt. Soc. Am. A*, **5**, 1127–1136.
- Horn, B. K. P. (1987). *J. Opt. Soc. Am. A*, **4**, 629–642.
- Huang, T. S., Blostein, S. D. & Margerum, E. A. (1986). *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* pp. 24–26. IEEE Computer Society.
- Huggins, D. J. (2014a). *J. Comput. Chem.* **35**, 377–385.
- Huggins, D. J. (2014b). *J. Chem. Theory Comput.* **10**, 3617–3625.
- Huynh, D. Q. (2009). *J. Math. Imaging Vis.* **35**, 155–164.
- Jupp, P. & Kent, J. (1987). *Appl. Stat.* **36**, 34–46.
- Kabsch, W. (1976). *Acta Cryst.* **A32**, 922–923.
- Kabsch, W. (1978). *Acta Cryst.* **A34**, 827–828.
- Karcher, H. (1977). *Comm. Pure Appl. Math.* **30**, 509–541.
- Kearsley, S. (1990). *J. Comput. Chem.* **11**, 1187–1192.
- Kearsley, S. K. (1989). *Acta Cryst.* **A45**, 208–210.
- Kneller, G. R. (1991). *Mol. Simul.* **7**, 113–119.
- Lesk, A. M. (1986). *Acta Cryst.* **A42**, 110–113.
- Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J. & Fulk, D. (2000). *Proc. ACM SIGGRAPH 2000*, pp. 131–144.
- Liu, P., Agrafiotis, D. K. & Theobald, D. L. (2010). *J. Comput. Chem.* **31**, 1561–1563.
- McLachlan, A. D. (1982). *Acta Cryst.* **A38**, 871–873.
- Manton, J. (2004). *Proc. 8th Int. Conf. Control Autom. Robot. Vis. (ICARCV 2004)*, Vol. 3, pp. 2211–2216. IEEE.
- Markley, F. L. (1988). *J. Astronaut. Sci.* **38**, 245–258.
- Markley, F. L., Cheng, Y., Crassidis, J. L. & Oshman, Y. (2007). *J. Guid. Contr. Dyn.* **30**, 1193–1197.
- Markley, F. L. & Mortari, D. (2000). *J. Astronaut. Sci.* **48**, 359–380.
- Moakher, M. (2002). *SIAM J. Matrix Anal. Appl.* **24**, 1–16.
- Nickalls, R. (1993). *Math. Gaz.* **77**, 354–359.
- Nickalls, R. (2009). *Math. Gaz.* **93**, 66–75.
- Nüchter, A., Lingemann, K., Hertzberg, J. & Surmann, H. (2007). *J. Field Robot.*, **24**, 699–722.
- Park, F. C. & Ravani, B. (1997). *ACM Trans. Graph.* **16**, 277–295.
- Rusinkiewicz, S. & Levoy, M. (2001). *Proc. Third Int. Conf. 3-D Digital Imaging Model.*, pp. 145–152. IEEE.
- Sarlette, A. & Sepulchre, R. (2009). *SIAM J. Contr. Optim.* **48**, 56–76.
- Schönemann, P. (1966). *Psychometrika*, **31**, 1–10.
- Shepperd, S. W. (1978). *J. Guid. Contr.* **1**, 223–224.
- Shoemake, K. (1985). *Comput. Graph.* **19**, 245–254.
- Shuster, M. D. & Natanson, G. A. (1993). *J. Astronaut. Sci.* **41**, 545–556.
- Theobald, D. L. (2005). *Acta Cryst.* **A61**, 478–480.
- Umeyama, S. (1991). *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 376–380.
- Wahba, G. (1965). *SIAM Rev.* **7**, 409.
- Walker, M. W., Shao, L. & Volz, R. A. (1991). *CVGIP Image Underst.* **54**, 358–367.
- Weisstein, E. W. (2019a). *Cubic formula*, <http://mathworld.wolfram.com/CubicFormula.html>.
- Weisstein, E. W. (2019b). *Quartic equation*, <http://mathworld.wolfram.com/QuarticEquation.html>.
- Wikipedia (2018a). *Kabsch algorithm*, <https://w.wiki/MoZ>.
- Wikipedia (2018b). *Wahba's problem*, <https://w.wiki/MR4>.
- Wikipedia (2019). *Ars Magna (Gerolamo Cardano)*, <https://w.wiki/Mob>.
- Zhang, Z. (2000). *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 1330–1334.