Noname manuscript No.
(will be inserted by the editor)

Rotation Averaging

Richard Hartley, Jochen Trumpf, Yuchao Dai, Hongdong Li

Received: date / Accepted: date

Abstract This paper is conceived as a tutorial on rotation averaging, summarizing the research that has been carried out in this area; it discusses methods for single-view and multiple-view rotation averaging, as well as providing proofs of convergence and convexity in many cases. However, at the same time it contains many new results, which were developed to fill gaps in knowledge, answering fundamental questions such as radius of convergence of the algorithms, and existence of local minima. These matters, or even proofs of correctness have in many cases not been considered in the Computer Vision literature.

We consider three main problems: single rotation averaging, in which a single rotation is computed starting from several measurements; multiple-rotation averaging, in which absolute orientations are computed from several relative orientation measurements; and conjugate rotation averaging, which relates a pair of coordinate frames. This last is related to the hand-eye coordination problem and to multiple-camera calibration.

Keywords geodesic distance · angular distance · chordal distance · quaternion distance · L_1 mean · L_2 mean · conjugate rotation

1 Introduction

In this paper, we will be interested in three different rotation averaging problems. In the following description, $d(\mathtt{R}, \mathtt{S})$ denotes the distance between two rotations R and S. Various different possible distance functions will be described later in the paper; for now, $d(\cdot, \cdot)$ is thought of as being any arbitrary metric on the space of rotations $\mathrm{SO}(3)$.

Address(es) of author(s) should be given

Single rotation averaging. In the single rotation averaging problem, several estimates are obtained of a single rotation, which are then averaged to give the best estimate. This may be thought of as finding a mean of several points $\mathtt{R}_i$ in the rotation space $\mathrm{SO}(3)$ (the group of all 3-dimensional rotations) and is an instance of finding a mean in a manifold.

Given an exponent $p \geq 1$ and a set of $n \geq 1$ rotations $\{\mathtt{R}_1, \ldots, \mathtt{R}_n\} \subset \mathrm{SO}(3)$ we wish to find the L^p-mean rotation with respect to d which is defined as

$$d^{\mathrm{p}}\text{-mean}(\{\mathtt{R}_1, \ldots, \mathtt{R}_n\}) = \operatorname*{argmin}_{\mathtt{R} \in \mathrm{SO}(3)} \sum_{i=1}^{n} d(\mathtt{R}_i, \mathtt{R})^p .$$

Since $\mathrm{SO}(3)$ is compact, a minimum will exist as long as the distance function is continuous (which any sensible distance function is). This problem has been much studied in the literature, but there are still open problems, some of which are resolved here.

Conjugate rotation averaging. In the conjugate rotation averaging problem, $n \geq 1$ rotation pairs $(\mathtt{L}_i, \mathtt{R}_i)$ (the left and right rotations) are given, and we need to find a rotation S such that $\mathtt{R}_i = \mathtt{S}^{-1}\mathtt{L}_i\mathtt{S}$ for all i. This problem arises when the rotations $\mathtt{R}_i$ and $\mathtt{L}_i$ are measured in different coordinate frames, and the coordinate transformation S that relates these two frames is to be determined.

In the presence of noise, the appropriate minimization problem is then to find

$$\operatorname*{argmin}_{\mathtt{S}} \sum_{i=1}^{n} d(\mathtt{R}_i, \mathtt{S}^{-1}\mathtt{L}_i\mathtt{S})^p .$$

This problem is sometimes referred to as the *hand-eye coordination problem*, see e.g. [13, 64, 84].

In the case where the individual rotations $\mathtt{R}_i$ and $\mathtt{L}_i$ are themselves estimated from relative orientation measurements $\mathtt{R}_{ij}$ and $\mathtt{L}_{ij}$, the two problems can be solved simultaneously to find S at the same time as the rotations $(\mathtt{R}_i, \mathtt{L}_i)$.

Multiple rotation averaging. In the multiple rotation averaging problem, several relative rotations $\mathtt{R}_{ij}$ are given, perhaps relating different coordinate frames, and n absolute rotations $\mathtt{R}_i$ are computed to satisfy the compatibility constraint $\mathtt{R}_{ij}\mathtt{R}_i = \mathtt{R}_j$. Only some $\mathtt{R}_{ij}$ are given, represented by index pairs (i,j) in a set $\mathcal{N}$. In the presence of noise, the appropriate minimization problem is expressed as seeking

$$\operatorname*{argmin}_{\mathtt{R}_1,\ldots,\mathtt{R}_n} \sum_{(i,j)\in\mathcal{N}} d(\mathtt{R}_{ij}, \mathtt{R}_j\mathtt{R}_i^{-1})^p \ .$$

For all these problems, we are interested in finding provably optimal and convergent solutions, mainly for the cases $p = 1$ and $p = 2$. This includes most particularly identifying the conditions under which the problems will allow a solution.

Our task in this paper is to report the known results about these problems, while at the same time filling in gaps of knowledge, particularly related to convergence, convexity or uniqueness of solutions to these problems.

Applications. The single-rotation averaging problem can be used in the case where several measurements of a single rotation $\mathtt{R}$ are given. These may be for instance measurements of the orientation of an object, derived from measurements taken with different cameras in a calibrated network. If the measurements are noisy, they can be averaged to find a mean. In another example, given a pair of images, several minimal sets of points (5 points for calibrated cameras) may be chosen and used to compute the relative rotation between the cameras. By a process of averaging, one may obtain the mean of these measurements, which provides an estimate of the true rotation relating the two cameras.

The multiple-rotation averaging problem has wide application to the problem of structure-from-motion (SfM), and several papers [55,73,29,37,71,41,38] have explored this method, often starting with an assumption that the rotations of the cameras are known. These rotations may be estimated separately by rotation averaging. This idea has been developed into a unified approach to SfM by Govindu [23,22,24], who also developed various rotation-averaging algorithms.

Conjugate rotation averaging is related to the hand-eye coordination problem, common in robotics [13,64,84]. In one formulation of this problem, consider a robot manipulating some object, which is also observed by a stationary camera. The orientation of the object can be computed at each moment through knowledge of the geometry of the robot (for instance, joint-angles). At the same time, the orientation of the object can be computed from the images taken from the camera. This gives two separate estimates of the orientation of the object (expressed as a rotation), but these are in different coordinate frames. By solving the conjugate rotation problem, one can compute the relationship between the robot and camera frames.

In another application, camera rigs used in robotic or mapping applications can consist of fixed cameras often with small or no overlap of fields of view. From SfM techniques, the trajectory of each camera may be computed independently. In the two-camera case this leads to pairs of rotations $(\mathtt{L}_i, \mathtt{R}_i)$. By solving the conjugate averaging problem, one may compute the relative orientation of the two cameras. This technique generalizes easily to several cameras. For best results, the conjugate averaging problem is solved simultaneously with the multiple-rotation averaging problem of determining the $\mathtt{R}_i$ and $\mathtt{L}_i$ [12].

Different metrics. Although the rotation averaging problem has been discussed frequently in the literature of Computer Vision, there has rarely been any discussion of what cost-function is actually being minimized by the algorithms in question. Discussion of this question in papers about optimization on manifolds has usually been more specific in this regard. The most common approach to the single-averaging problem is to find the Karcher mean [26,40] which is defined as

$$\mathbf{y}^* = \operatorname*{argmin}_y \sum_{i=1}^n d_{\text{geod}}(\mathbf{x}_i, \mathbf{y})^2 \tag{1}$$

where $\mathbf{x}_i; i = 1, \ldots, n$ are several points on a Riemannian manifold, and $d_{\text{geod}}(\cdot,\cdot)$ represents the minimal geodesic distance between two points. The choice of the squared-distance in this expression means that we are minimizing a least-squares (L_2) cost function. This definition is easily generalized to include other than sum of squares costs. The most immediate generalization is to minimize the L_1 cost, namely the sum of (unsquared) distances $d_{\text{geod}}(\mathbf{x}_i, \mathbf{y})$. We will refer to this as the geodesic L_1-mean of the points. Other exponents, such as $d_{\text{geod}}(\mathbf{x}_i, \mathbf{y})^q$ are possible, but will not be considered in any detail in this paper. Thus, by referring to a geodesic mean, we imply the minimization of a cost based on geodesic distance in the manifold itself. The literature on the Karcher mean is very large, see e.g. [26,40,11, 47,2] and the references therein. Papers relating to computation of the Karcher mean for rotations include [57], [48], [53] and [45], with Manton [53] giving a simple iterative solution.

Computation of the geodesic L_1-mean in a manifold has received much less attention. Recent work includes L_1 minimization on $\mathrm{SO}(3)$ [12], which suggests a gradient-descent algorithm. This problem has been solved in the more general context of a Riemannian manifold with positive sectional curvature in [20] and extended in [83]. The solution of [20] involves iterative steps of the Weiszfeld algorithm [81] in tangent spaces of the manifold. This literature will be surveyed in more detail later.

In the context of rotations in $\mathrm{SO}(3)$, the (natural) geodesic metric $d_{\text{geod}}(\cdot,\cdot)$ is equal to the angle between two rotations. Specifically, given rotations $\mathtt{R}$ and $\mathtt{S}$, the product $\mathtt{RS}^{-1}$ is also a rotation, about some axis by an angle θ in the range $0 \leq \theta \leq \pi$. We define $d_\angle(\mathtt{R}, \mathtt{S}) = \theta$, and refer to it as the angle metric. It will turn out that this is identical with the geodesic metric on $\mathrm{SO}(3)$, so we will sometimes also refer to it as the geodesic metric.

Other metrics exist, other than the geodesic metric. The so-called "chordal" metrics relate to a specific embedding of a manifold in a Euclidean space $\mathbb{R}^N$. The distance between two points in the manifold is then defined to be the Euclidean distance in $\mathbb{R}^N$ between the embedded points. A rotation $\mathtt{R} \in \mathrm{SO}(3)$ is commonly represented by a 3×3 orthogonal matrix (with unit determinant). There is therefore a natural embedding of a rotation $\mathtt{R}$ in $\mathbb{R}^9$. Given two rotations $\mathtt{R}$ and $\mathtt{S}$, their chordal distance is then the distance between their embeddings in $\mathbb{R}^9$. This is equal to $d_{\mathrm{chord}}(\mathtt{R}, \mathtt{S}) = \|\mathtt{R} - \mathtt{S}\|_{\mathrm{F}}$, where $\| \cdot \|_{\mathrm{F}}$ is the Frobenius norm of the matrix. It will be shown later that $d_{\mathrm{chord}}(\mathtt{R}, \mathtt{S}) = 2\sqrt{2}\,\sin(\theta/2)$, where $\theta = d_\angle(\mathtt{R}, \mathtt{S})$.

A further representation of rotations as points in a Euclidean space is through quaternions, in which rotations are represented as unit 4-vectors. This allows us to define another "chordal" distance between rotations equal to the distance between their quaternion representations. However, since a given quaternion and its negative both represent the same rotation, we define the minimum of the two possible distances between $\pm\mathbf{r}$ and $\pm\mathbf{s}$ to be the quaternion distance $d_{\mathrm{quat}}(\mathtt{R}, \mathtt{S})$ between the corresponding rotations. It will be shown later that $d_{\mathrm{quat}}(\mathtt{R}, \mathtt{S}) = 2\sin(\theta/4)$.

The reason for considering different metrics on $\mathrm{SO}(3)$ as a basis for averaging is that certain known simple algorithms naturally minimize cost functions involving chordal or quaternion distance. From the point of view of understanding the algorithms, it is essential to understand what metric is being minimized.

Approach and Prerequisites. Rotation space $\mathrm{SO}(3)$ naturally forms a Lie Group, an algebraic group with a manifold structure. It consequently also has the structure of a Riemannian manifold. It is natural to use the language of Lie groups, Lie algebras, Riemannian metrics, geodesics, tangent spaces, exponential maps, and all the machinery of Riemannian and differential manifolds when discussing $\mathrm{SO}(3)$. In this paper, although these terms will be used at times as a convenient descriptive language, there will be no appeal to any advanced concepts related to Riemannian manifolds or Lie Groups. An effort has been made to present the material in a way that requires only relatively elementary mathematical concepts, and when more advanced concepts are used (for example concepts from manifold topology such as fundamental groups or covering spaces), they are motivated by intuitive descriptions. For instance, geodesics are defined simply to be locally shortest paths on a manifold; all the required properties are derived using elementary concepts.

Since the word "manifold" itself is often used in Computer Vision in a somewhat loose sense, it bears stating that the word is used in this paper in its strict mathematical sense of a locally Euclidean Hausdorff space whose topology has a countable base.[1] 'Locally Euclidean' just means that each point has some neighbourhood that is homeomorphic to an open ball in $\mathbb{R}^N$ for some N. In the case of $\mathrm{SO}(3)$, the dimension $N = 3$, so $\mathrm{SO}(3)$ is a 3-manifold.

New Results. Although this paper aims at summarizing the state of knowledge in rotation averaging, it does contain several results that were previously unknown, or unproven. Here, we enumerate the major new results of this paper. Note that some of these results were previously announced in our recent conference papers [12,31,27].

1. The recognition of the role of *weakly convex sets* (definition 1) in the analysis of convexity of distance metrics on $\mathrm{SO}(3)$ is new. Their characterization (theorem 10) has not been previously known; most importantly, the systematic study of the region of convexity of the given distance metrics on $\mathrm{SO}(3)$ (theorem 3) significantly extends previously known results since it is based on the notion of weak convexity where previous results were based on the much stronger notion of (geodesic) convexity. See also [31].
2. The proof that any global minimum of the single rotation averaging cost function for points in a convex set must also lie in the convex set (theorem 5) is stated for the first time explicitly for $\mathrm{SO}(3)$. A similar result has been shown in the more general context of Riemannian manifolds, but under more restrictive conditions on the size of the convex set in [2]. See also [31].
3. The analysis of the multiple rotation quaternion averaging algorithm [22] is new (section 7.1).
4. The proof of existence of local minima of the multiple rotation averaging cost function with cost close to the global minimum (section 7.3) is new.

Structure of the paper. Following this introduction, the next section summarises previous work on rotation averaging in computer vision, robotics, structural chemistry and other related areas. Section 3 contains a detailed discussion of representations for rotations and their mutual relationships, including orthogonal matrix representations, angle-axis representations, unit quaternions and several others. Section 4 introduces the distance measures we consider in this paper and discusses geodesics (locally shortest paths)

[1] "La notion générale de variété est assez dificile à définir avec précision. [The general notion of a manifold is rather difficult to define with precision.]" [9, page 56]

with respect to these metrics. We also state and prove a general cosine rule for $\mathrm{SO}(3)$. Sections 5, 6 and 7 contain the main results on the single, conjugate and multiple rotation averaging problems, respectively. The paper concludes with two appendices. The first appendix contains a detailed discussion of convexity in $\mathrm{SO}(3)$. We introduce the notion of weak convexity and discuss the convexity properties of the various distance metrics in $\mathrm{SO}(3)$. The second appendix provides formulas for gradients and Hessians of the averaging cost functions on $\mathrm{SO}(3)$.

2 Previous Work on Rotation Averaging

The rotation averaging problem arises frequently in many research areas ranging from pure fundamental mathematical exploration to practical engineering and scientific applications, such as computer vision, robotics and structural chemistry.

2.1 Rotation averaging in robotics

Most applications in robotics involve the full special euclidean group $\mathrm{SE}(3)$, a semidirect product of the rotation group $\mathrm{SO}(3)$ with the (additive) group $\mathbb{R}^3$ of translations. Elements of $\mathrm{SE}(3)$ are used to encode the "pose" of a robot in its 3D environment where pose comprises both "orientation" or "attitude" (the rotation part) as well as "position" or "location" (the translational part) with respect to a fixed reference frame.

A Consistent Pose Registration (CPR) framework was proposed by Lu and Milios [52] for the task of mobile robot Simultaneous Localization and Mapping (SLAM), in which a globally consistent configuration of the robot's poses at different times is built by fusing (averaging) all local relative poses. However, Lu and Milios' work is confined to the case of 3 degrees of freedom planar motion which is substantially simpler than the 6 degrees of freedom case where our work could be applied, because in the planar motion case two rotations about the same point always commute. This is not the case for 3D rotations. Agrawal [3] presented a Lie algebraic approach for consistent pose registration for general Euclidean motion.

The hand-eye coordination problem is the same as our conjugate rotation averaging problem and has been discussed extensively [13, 64, 84]. In these papers, no optimality is shown nor is it shown what objective function, in terms of what metric, is being minimzed. Strobl and Hirzinger [75] approached the problem by defining a metric on the group $\mathrm{SE}(3)$ and a corresponding error model for nonlinear optimization. The metric for rotation error is given as a weighted version of the rotation angle.

2.2 Rotation averaging in computer vision

Structure from Motion. In computer vision and multi-view geometry, Govindu seems to be the first who introduced the idea of motion averaging for structure-from-motion computation. He published a series of papers addressing this problem [22–24]. In [22] a simple linear least squares method is proposed where rotations in $\mathrm{SO}(3)$ are parametrized by quaternions and a closed-form linear least squares solution is derived, using the Singular Value Decomposition (SVD). The paper [23] follows a nonlinear optimization on manifold approach which is similar in spirit to the algorithms we discuss here. Another paper by Govindu [24] tackles robustness problems by adopting a RANSAC-type approach for outlier-removal.

Martinec and Pajdla [55] discussed rotation averaging using the chordal metric in $\mathbb{R}^{3\times3}$ and compared their method with the linear quaternion method. This approach to averaging using the chordal metric has similar problems as linear quaternion averaging. The obtained result is not necessarily a proper rotation before manifold projection is performed.

Gramkow [25] compared three different methods for single rotation averaging, that is, from orthogonal rotation matrices, from unit quaternion representations and from angle-axis representations, and showed that the results are quite similar if the individual rotations are close enough. In our present paper, we also consider the three cases (we call them the chordal metric, the quaternion metric, and the angle metric respectively), and provide rigorous theoretic analysis and detailed algorithm implementations.

When covariance uncertainty information is available for each local measurement, Agrawal shows how to incorporate such information in the Lie-group averaging computation [3]. Alternatively, one could apply the belief propagation framework to take the covariance information into account [14].

Calibration. Often several cameras are attached rigidly to a platform, such as a moving vehicle, and used to capture large amounts of video. In analyzing such imagery, it is possible to consider several cameras as a single "generalized" camera [66, 6]. To be able to do this, however, it is necessary to calibrate the set of cameras. In particular, this means that the relative placement of all the cameras must be determined.

Non-overlapping multi-camera rigs are of particular interest in practice. As the component cameras have little or no overlap in their fields of views, the effective overall field of view is wider, leading to efficient data acquisition. However, because of the non-overlap, calibration is a potential problem, which has been considered in several papers.

Calibration using mirrors has been frequently suggested [76,46,70]. Methods that simply use the image-tracks from each camera separately have also been proposed [17,42,10, 43,50,44,12,49]. This is an instance of the conjugate rotation averaging problem discussed in this paper. The sequence of orientations of each camera in its own frame, may be computed from the sequence of images taken by that camera. Subsequently, the conjugate rotation problem is used to determine the relative orientations of all the cameras.

Consensus rotation averaging in distributed camera networks. Recent developments in wireless sensor network technology have led to the deployment of distributed camera networks where camera and processing nodes may be spread over a geographical area, with no centralized processing unit and limited ability to communicate large amounts of information over long distances. These networks require new techniques for calibrating camera networks and structure from motion.

Most computer vision algorithms assume that all the data (the images) are available on a single computer where centralized processing is possible. However, this paradigm is inherently incompatible with sensor networks for two reasons. Firstly, it requires the transmission of large volumes of raw data. Secondly, it demands processing resources not available in mote-class devices. A multiple rotation averaging algorithm can be applied naturally to a distributed camera network as it is a local averaging algorithm involving only the neighbouring camera nodes. Through iteration, each camera will obtain its pose (both rotation and translation) in the global coordinate system.

To process video data on distributed nodes, the camera network must be accurately calibrated in both space and time [68]. In distributed camera network applications, Lie-averaging techniques have been applied to the distributed calibration of a camera network [79]. Antone and Teller [78] considered calibration of a number of unordered views by fusing rotations via a visibility graph.

2.3 Rotation averaging in structural chemistry

In structural chemistry (e.g. the computation of crystal structures), it is often of interest to analyze grain orientations in polycrystalline material, which sometimes requires the computation of the mean orientation. Humbert et al [34,35] proposed two methods (quaternion and rotation matrix averaging) for such a task. A variant of the quaternion algorithm using 4×4 eigendecomposition was given by Morawiec [59]. Morawiec [58] also pointed out some theoretical inaccuracies in Humbert's two original algorithms, including the sign ambiguity associated with the quaternion representation. For a complete treatment of this topic in the crystallography field, the reader is referred to a recent monograph by Morawiec [60].

2.4 Other related research

A general mathematical exposition of the single rotation averaging problem can be found in [57], where several different definitions of mean rotation are given under different metrics. Pennec [65] provided a thorough discussion of stochastic "mean objects" on homogeneous Riemannian manifolds. The obtained geometric mean depends only on intrinsic characteristics of the manifold in question. This work ties in with the previously mentioned large body of work on the Karcher mean, see [26,40,11,47,2] and the references therein (cf. section 1). Pennec suggested a gradient descent algorithm to compute mean rotations, see also [57, 48,53,45]. Besides the simple least squares mean, Pennec also studied weighted least squares means and the Riemannian Mahalanobis mean based on predicted uncertainty covariance at the estimated mean object.

Quaternion averaging was studied in some detail by Markley et al. [54], who were motivated by a problem in aerospace engineering, namely spacecraft attitude estimation from multiple star trackers.

Buchholz and Sommer [8] describe how to compute means on Clifford groups, a problem that can be viewed as a generalization of quaternion averaging, allowing a general treatment of approximated averaging for all classical groups. Fiori and Tanaka [18] introduced a novel procedure for designing an averaging algorithm for a committee of learning machines under the assumption that the machines share a common parameter space, namely the group $\mathrm{SO}(p)$ of special orthogonal matrices. Sarlette and Sepulchre [72] formulated consensus as an optimization problem and designed distributed consensus algorithms for N agents moving on a connected compact homogeneous manifold.

The problem of finding the L_1-mean of a set of points in $\mathbb{R}^N$ for $N > 1$ is a classical problem, going back at least to Fermat. The special case of this problem for three points forming a triangle in $\mathbb{R}^2$ was solved by Torricelli. The solution is the so-called Fermat point of the triangle, provided no angle exceeds $120°$. The problem subsequently was studied in some detail by Weber [80]. For this reason, it is sometimes referred to as the Fermat-Weber problem or simply the Weber problem. It is also called the "location" problem. This latter name is related to its interpretation in terms of optimal placement of a factory to minimize the sum of its distances to a set of resources. The solution is commonly referred to as the geometric median of the points. The Weiszfeld algorithm [81] is a well-known algorithm for finding the L_1 mean of a set of points in $\mathbb{R}^n$. Refinements to the basic algorithm include geometric speed-up methods [63] and Newton methods [51]. However, the simplicity of the basic Weiszfeld algorithm and the rapidity with which the upgrade may be computed make it a very attractive algorithm even when compared to its more sophisticated versions. The Weiszfeld algorithm may also be generalized to Banach spaces [15] and to Riemannian manifolds [20,83]. This last case is of relevance to the problem of computing the L_1 geodesic mean on $\mathrm{SO}(3)$ [27].

Other problems that are closely related to the single rotation-averaging problem are also investigated by computer vision researchers. These include:

1. Principal Component Analysis on manifolds [19];
2. Nonlinear mean-shift on Riemannian manifolds [77];
3. Geodesic k-means clustering [5].

3 Alternative Pictures of Rotation Space

We begin by discussing several different representations of the set $\mathrm{SO}(3)$ of all rotations of 3-dimensional Euclidean space. While each of these representations is well discussed and often used in the literature, we find that none of them is universally suitable for the discussion of all aspects of all the problems we cover in this paper. We briefly review these different geometric pictures.

Throughout this paper we will use the language of Lie groups and (occasionally) Lie algebras, but our development will be self-contained, and will not rely on anything other than elementary knowledge of the theory of Lie groups. A Lie group is a group G which is at the same time a differentiable manifold having the property that a mapping $G \to G$ induced by left or right multiplication by a fixed element $g \in G$ is smooth, and the mapping $g \mapsto g^{-1}$ is smooth.

For a more in-depth discussion of the use of group theory in computer vision see the book by Kanatani [39].

3.1 The Matrix Lie Group $\mathrm{SO}(3)$

The set of rotations

$$\mathrm{SO}(3) = \{\mathtt{R} \in \mathbb{R}^{3\times 3} \,|\, \mathtt{R}^\top \mathtt{R} = \mathtt{I}_{3\times 3},\, \det(\mathtt{R}) = 1\}$$

forms a matrix Lie group, a subgroup of the general linear group $\mathrm{GL}(3)$ of invertible 3×3-matrices, namely the orthogonal matrices $\mathtt{R}$ with $\det \mathtt{R} = 1$.

Associated with the Lie group $\mathrm{SO}(3)$ is the Lie algebra $\mathfrak{so}(3)$ consisting of the set of all skew-symmetric 3×3-matrices. The connection between these two entities is the *exponential map* taking an element $\Omega \in \mathfrak{so}(3)$ to its matrix exponential $\exp(\Omega)$ which is an element in $\mathrm{SO}(3)$. In fact, any rotation $\mathtt{R} \in \mathrm{SO}(3)$ may be expressed in the form

$$\mathtt{R} = \exp(\Omega) = \mathtt{I} + \Omega + \Omega^2/2! + \Omega^3/3! + \ldots$$

where Ω is a 3×3 skew-symmetric matrix; the exponential map is surjective, onto $\mathrm{SO}(3)$. It is also locally one-to-one.

A matrix Ω may be represented in terms of the entries of a 3-vector $\mathbf{v} = (v_1, v_2, v_3)^\top$ by

$$\Omega = [\mathbf{v}]_\times = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix} \tag{2}$$

so the skew-symmetric 3×3 matrices form a vector space isomorphic to $\mathbb{R}^3$. It follows from these remarks that the Lie group $\mathrm{SO}(3)$ is a manifold of dimension 3, embedded in $\mathbb{R}^{3\times 3}$.

By referring to $\mathfrak{so}(3)$ as a Lie algebra, we imply the existence of a Lie-bracket operation. This is the matrix commutator $[\Omega, \Gamma] = \Omega\Gamma - \Gamma\Omega$, but we will make little use of this concept.

3.2 The Angle-Axis Representation

Every rotation in $\mathrm{SO}(3)$ can also be represented as a rotation through an angle θ about an axis represented by a unit 3-vector $\hat{\mathbf{v}}$. The vector $\mathbf{v} = \theta\hat{\mathbf{v}}$ is known as the angle-axis representation of the rotation. Note that by this definition, the angle-axis representation is not unique, since an alternative representation is $(2\pi - \theta)(-\hat{\mathbf{v}})$. The connection between the angle-axis representation of a rotation and its 3×3 matrix representation is as follows. Given a 3-vector $\mathbf{v} = \theta\hat{\mathbf{v}}$, it is shown (for instance) in [32] that the matrix $\exp[\mathbf{v}]_\times$ is precisely the rotation through angle θ about the axis represented by the unit vector $\hat{\mathbf{v}}$. Thus, the mapping $\exp[\cdot]_\times$ from $\mathbb{R}^3$ to $\mathrm{SO}(3)$ connects the two representations of a rotation.

Every rotation can be represented as a rotation through some angle by at most π radians. In fact, if the rotation is by less than π radians, the representation is unique. A rotation through angle π about an axis $\hat{\mathbf{v}}$ may equally well be represented as a rotation through π about the oppositely-oriented axis $-\hat{\mathbf{v}}$. Thus, the mapping $\exp[\cdot]_\times$ is surjective, and is one-to-one on the open ball in $\mathbb{R}^3$ of radius π. The mapping is two-to-one on the boundary of this ball. In this way, we may think of rotation space as being represented by the closed ball $B_\pi \subset \mathbb{R}^3$ with opposite points on its boundary identified. By identifying opposite points on the boundary of a closed ball in $\mathbb{R}^3$, we obtain [56] the projective space $\mathbb{P}^3$. Hence, topologically $\mathrm{SO}(3)$ is homeomorphic to $\mathbb{P}^3$.

Since we will frequently be concerned with this correspondence between the angle-axis representation and the matrix representation of rotations, we adopt a minor abuse of terminology by referring to the mapping $\exp[\cdot]_\times : \mathbb{R}^3 \to \mathrm{SO}(3)$ as the *exponential map* and its inverse as the *logarithm map*, $\log(\cdot) : \mathrm{SO}(3) \to \mathbb{R}^3$. This terminology is justified if we look upon $\mathbb{R}^3$ as the tangent space to $\mathrm{SO}(3)$ at the identity. Since the exponential map is not one-to-one, its inverse is not strictly defined. We resolve this by defining $\log(\mathtt{R})$ to be the angle-axis vector of length no more than π, which is uniquely defined unless R is a rotation through π radians, in which case we let $\log(\mathtt{R})$ be one of the two possible vectors of length π representing this rotation. The angle of rotation of R is hence equal to $\|\log \mathtt{R}\|_2$ where the norm is the Euclidean norm in $\mathbb{R}^3$.

Considering now the Lie-algebra, we observe that the mapping $[\cdot]_\times : \mathbb{R}^3 \to \mathfrak{so}(3)$ is a vector space isomorphism (it preserves addition). Moreover, if we define a Lie-bracket operation $\mathbb{R}^3$ by the vector product $[\mathbf{v}, \mathbf{w}] = \mathbf{v} \times \mathbf{w}$, then this map is a Lie-algebra isomorphism between $\mathbb{R}^3$ and $\mathfrak{so}(3)$, where the Lie-bracket operation on $\mathfrak{so}(3)$ was defined above by the commutator.

The exponential map on $\exp[\cdot]_\times : \mathbb{R}^3 \to \mathrm{SO}(3)$ can be computed using Rodrigues' formula (see for instance [32]):

$$\exp(\theta\hat{\mathbf{v}}) = \mathtt{I} + \sin(\theta)[\hat{\mathbf{v}}]_\times + (1 - \cos(\theta))([\hat{\mathbf{v}}]_\times)^2 \ . \qquad (3)$$

The logarithm can be computed using the formula

$$\log(\mathtt{R}) = \begin{cases} \arcsin(\|\mathbf{y}\|_2)\dfrac{\mathbf{y}}{\|\mathbf{y}\|_2}, & \mathbf{y} \neq \mathbf{0} \\ \mathbf{0}, & \mathbf{y} = \mathbf{0} \end{cases}$$

where $\mathbf{y} = (y_1, y_2, y_3)$ is computed from

$$\frac{1}{2}(\mathtt{R} - \mathtt{R}^\top) = \begin{pmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{pmatrix} .$$

3.3 The Quaternion Sphere

"Anyone who has ever used any other parametrization of the rotation group will, within hours of taking up the quaternion parametrization, lament his or her misspent youth." [4]

The group of quaternions is of fundamental importance in the study of rotations. This group consists of the set of non-zero real 4-vectors $\mathbb{R}^4$, equipped with a multiplication defined as follows. Let $\mathbf{r}_1 = (c_1, \mathbf{v}_1)$ and $\mathbf{r}_2 = (c_2, \mathbf{v}_2)$ be two quaternions, where $\mathbf{v}_i$ is the vector made up of the last three components of the quaternion. Multiplication is defined by

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = (c_1c_2 - \langle \mathbf{v}_1,\, \mathbf{v}_2\rangle \,, c_1\mathbf{v}_2 + c_2\mathbf{v}_1 + \mathbf{v}_1 \times \mathbf{v}_2) \ .$$

Here, $\langle \mathbf{v}_1,\, \mathbf{v}_2\rangle$ is the standard inner product and $\times$ represents the vector or "cross" product of the 3-vectors involved. Another way to formulate the multiplication operation is to represent a quaternion $\mathbf{r} = (r_0, r_1, r_2, r_3)$ by writing $\mathbf{r} = r_0 + r_1\mathbf{i} + r_2\mathbf{j} + r_3\mathbf{k}$, where r_0 is thought of as the *real part* of the quaternion, and i, j and k are purely *imaginary* components. Multiplication of two quaternions $(r_0 + r_1\mathbf{i} + r_2\mathbf{j} + r_3\mathbf{k}) \cdot (s_0 + s_1\mathbf{i} + s_2\mathbf{j} + s_3\mathbf{k})$ is carried out by applying the distributive law to multiply out the product, and using the identities

$$\mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = \mathbf{i} \cdot \mathbf{j} \cdot \mathbf{k} = -1 \ .$$

An important property of quaternion multiplication is that $\|\mathbf{q}_1 \cdot \mathbf{q}_2\| = \|\mathbf{q}_1\|\,\|\mathbf{q}_2\|$, where $\|\mathbf{r}\|$ represents the norm of the quaternion, equal to its Euclidean norm as a 4-vector. The non-zero quaternions form a group under this multiplication operation. The group identity is the quaternion $(1, 0, 0, 0)$, and the inverse of $\mathbf{r} = (c, \mathbf{v})$ is $\mathbf{r}^{-1} = (c, -\mathbf{v})/\|\mathbf{r}\|^2$. The unit length quaternions form a subgroup of the quaternion group.

With this defined multiplication, the unit quaternions evidently form a Lie group, being at the same time a group, and a smooth manifold of dimension 3. One of the properties of a Lie group is that the multiplication operation must be continuous. It is instructive to understand the global action of the multiplication operation. For a fixed unit quaternion $\mathbf{r}$, consider the map $\mathbf{q} \mapsto \mathbf{r} \cdot \mathbf{q}$. Since quaternion multiplication is verifiably bilinear in the entries of the quaternions, this mapping can be written in terms of a matrix-vector product as $\mathbf{q} \mapsto \mathtt{P}_\mathbf{r}\mathbf{q}$, where $\mathtt{P}_\mathbf{r}$ is a 4×4 matrix with entries determined by $\mathbf{r}$. In addition, since for all vectors $\mathbf{q}$, we have

$$\|\mathbf{q}\| = \|\mathbf{r} \cdot \mathbf{q}\| = \|\mathtt{P}_\mathbf{r}\mathbf{q}\|,$$

it follows that $\mathtt{P}_\mathbf{r}$ is an orthogonal matrix. Therefore, multiplication by $\mathbf{r}$ has the effect of applying an orthogonal transformation, or rotation, to the unit quaternion sphere.

Quaternions as rotations. A rotation R may be represented by a unit quaternion $\mathbf{r}$ as follows. If $\hat{\mathbf{v}}$ is the unit vector representing the axis of the rotation and θ is the angle of the rotation about that axis, then $\mathbf{r}$ is defined as

$$\mathbf{r} = (\cos(\theta/2), \hat{\mathbf{v}}\sin(\theta/2)). \qquad (4)$$

We write $\mathbf{r} \mapsto \mathtt{R}$ to indicate the mapping from the unit quaternions to $\mathrm{SO}(3)$ indicated by the inverse correspondence. This may be expressed formally as

$$\mathbf{r} = (\cos(\theta/2), \hat{\mathbf{v}}\sin(\theta/2)) \mapsto \exp[\theta\hat{\mathbf{v}}]_\times = \mathtt{R} \ .$$

This mapping preserves multiplication, in that if $\mathbf{r} \to \mathtt{R}$ and $\mathbf{s} \to \mathtt{S}$, then $\mathbf{r} \cdot \mathbf{s} \to \mathtt{RS}$. Thus, this mapping is a Lie group homomorphism in which quaternion multiplication corresponds to ordinary matrix multiplication of rotations.

Both $\mathbf{r}$ and $-\mathbf{r}$ represent the same rotation, that is, the homomorphism from the unit quaternions to $\mathrm{SO}(3)$ is a 2-to-1 mapping. Topologically, the unit quaternions form a unit sphere S^3 in $\mathbb{R}^4$, and there is a 2-to-1 mapping from S^3 onto $\mathrm{SO}(3)$ in which opposite points of the sphere are identified. This mapping is evidently continuous. In the language of topology, S^3 is a two-fold covering space (or double cover) of $\mathrm{SO}(3)$. If we restrict ourselves to rotations through angles less than π then these are in 1-to-1 correspondence to points of the upper quaternion hemisphere with the "north pole" $(1, 0, 0, 0)$ corresponding to the identity rotation (rotation through an angle of 0). In this picture, the "equator" of the quaternion sphere corresponds exactly to the rotations through an angle of π with opposite points on the equator representing the same rotation. This picture of $\mathrm{SO}(3)$ in which we picture rotations as points on the unit 3-sphere (with opposite points representing the same rotation) will be one of our most common ways of visualizing $\mathrm{SO}(3)$. Once more, this picture indicates that $\mathrm{SO}(3)$ is homeomorphic to projective 3-space $\mathbb{P}^3$.

Relation to the angle-axis formulation. The quaternion $\mathbf{q} = (\cos(\theta/2), \sin(\theta/2)\hat{\mathbf{v}}) = (c, \mathbf{v})$ represents a rotation about the unit axis $\hat{\mathbf{v}} = \mathbf{v}/\|\mathbf{v}\|$ through an angle $\theta = 2\arccos(c)$. Hence, we deduce that the angle-axis representation of the quaternion $\mathbf{q} = (c, \mathbf{v})$ is $2\arccos(c)\,\mathbf{v}/\|\mathbf{v}\|$, or alternatively $2\arccos(-c)\,(-\mathbf{v}/\|\mathbf{v}\|)$.

3.4 The Gnomonic Projection

Starting from the representation of $\mathrm{SO}(3)$ as the quaternion sphere, S^3 visualized as the unit sphere embedded in $\mathbb{R}^4$, the gnomonic projection of S^3 to $\mathbb{R}^3$ is the projection from the centre of the sphere, $(0, 0, 0, 0)$, onto a tangent (3-dimensional) hyper-plane. For simplicity, we may consider this to be the tangent hyper-plane passing through the point $(-1, 0, 0, 0)$ on S^3, that is, the "south pole", representing the identity rotation. Clearly, this is a 2-to-1 projection of S^3, since opposite points on the sphere project to the same point.

Since a great circle on S^3 is the intersection of S^3 with a (2-dimensional) plane passing through the centre point $(0, 0, 0, 0)$, namely the plane spanned by the radius vector of any point on the great circle and a tangent vector along the great circle at that point, we easily see that the projection of a great circle is the intersection between this plane and the projection hyper-plane. This shows that the projection of a great circle on S^3 is a straight line in the projection hyper-plane. This type of map is sometimes also called a Beltrami map [7] in the literature.

In S^3, the "equator" is the intersection of the "equatorial hyper-plane" consisting of points $(0, x, y, z)$, with the sphere. Projecting from the origin, we see that the equator maps to the "plane at infinity" in $\mathbb{R}^3$. More exactly, we see that the gnomonic projection maps S^3 to $\mathbb{R}^3 \cup \Pi_\infty$, which is a 3-dimensional projective space, topologically homeomorphic to $\mathrm{SO}(3)$. Geodesics in $\mathrm{SO}(3)$ correspond to straight-lines in $\mathbb{R}^3$ along with straight lines in the plane at infinity. We will see later that this representation of $\mathrm{SO}(3)$ is particularly useful when it comes to concepts like geodesics and convexity.

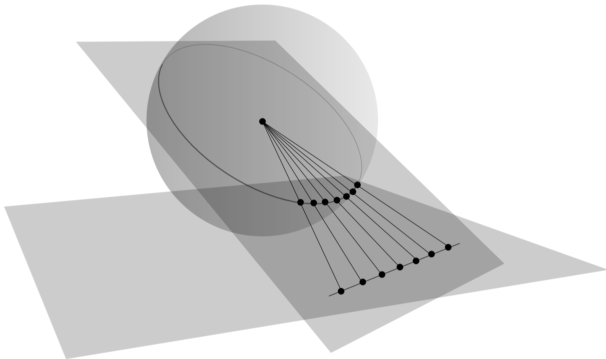


Fig. 1 Gnomonic projection of a sphere.

The above paragraphs described the gnomonic projection localized at the identity rotation, since the tangent hyper-plane was chosen to pass through a point in the quaternion sphere representing the identity rotation. One may equally well construct a gnomonic projection, with similar properties, localized about any other rotation (point on the quaternion sphere).

The parametrization of rotations through angles less than π given by the cartesian coordinates of the gnomonic projection of the upper quaternion hemisphere is usually called the Rodrigues parametrization, not to be confused with Rodrigues' formula (3). Assembling these Rodrigues parameters into a vector yields the so-called Gibbs vector associated with the rotation. The rotation axis $\hat{\mathbf{v}}$ is related to the Gibbs vector through a factor of $\tan(\theta/2)$, where θ is the rotation angle [60]. In other words, the Gibbs vector is equal to $\tan(\theta/2)\,\hat{\mathbf{v}}$. Table 3.4 shows three different vectorial parametrizations of a rotation.

Quaternion	$(\cos(\theta/2), \sin(\theta/2)\hat{\mathbf{v}})$
Angle-axis	$\theta\,\hat{\mathbf{v}}$
Gibbs/Rodrigues	$\tan(\theta/2)\,\hat{\mathbf{v}}$

Table 1 *Three different vectorial parametrizations for the rotation through angle θ about the unit axis $\hat{\mathbf{v}}$.*

The Cayley Transform. The Cayley transform on matrices is the mapping $\mathtt{A} \mapsto \mathtt{A}^c = (\mathtt{I} - \mathtt{A})\,(\mathtt{I} + \mathtt{A})^{-1}$, which is defined for any square matrix, provided that $(\mathtt{I} + \mathtt{A})$ is invertible. The Cayley transform is its own inverse, so $(\mathtt{A}^c)^c = \mathtt{A}$.

The relevance of the Cayley transform to rotations is as follows [82].

Proposition 1 *The Cayley transform of a rotation matrix* $\mathtt{R} \in \mathrm{SO}(3)$ *is a skew-symmetric matrix, and vice versa. Thus the correspondence* $\mathtt{R} \overset{c}{\leftrightarrow} [\mathbf{v}]_\times$ *is a one-to-one correspondence between skew-symmetric matrices and rotations* $\mathtt{R}$*, excluding rotations through an angle of* π.

The Cayley transform is closely related to the gnomonic projection, as follows. Applying the Cayley transform to a rotation, we obtain a skew-symmetric matrix $[\mathbf{v}]_\times$. This defines a correspondence $\mathtt{R} \leftrightarrow \mathbf{v}$ between rotations and 3-vectors. A simple calculation shows that applying the gnomonic projection to the quaternion $\mathbf{r}$ corresponding to $\mathtt{R}$ leads to the same vector $\mathbf{v}$. Thus, the Cayley transform and the gnomonic projection are essentially the same map, applied to the matrix and quaternion representations of a rotation. The Cayley transform is not defined for rotations through an angle of π (since $(\mathtt{I} + \mathtt{R})$ is then not invertible). Such rotations correspond to quaternions on the "equatorial plane", and hence to points at infinity under the gnomonic projection.

3.5 Projective Geometric Model

As we have discussed above, $\mathrm{SO}(3)$ is topologically equivalent to the 3-dimensional projective space, $\mathbb{P}^3$. In fact, the gnomonic projection maps the quaternion sphere (and hence $\mathrm{SO}(3)$) to $\mathbb{P}^3 = \mathbb{R}^3 \cup \Pi_\infty$, a standard model for the projective space. Note that the mapping from $\mathrm{SO}(3)$ to the quaternion sphere is a 1-to-2 mapping, since both a quaternion and its negative represent the same rotation. The gnomonic projection on the other hand maps opposite points on the sphere to the same point in $\mathbb{R}^3 \cup \Pi_\infty$, so the composite mapping is a one-to-one mapping from $\mathrm{SO}(3)$ onto $\mathbb{P}^3$.

In this mapping, as noted, great circles in the quaternion sphere map to the lines in $\mathbb{P}^3$. In addition, planes in $\mathbb{P}^3$ arise as the projection of "great" 2-spheres in the quaternion sphere. Choosing different tangent planes to the quaternion sphere on which to localize the gnomonic map is equivalent to choosing different planes in $\mathbb{P}^3$ to be the "plane at infinity."

The usual geometric model for the projective plane is the Euclidean space $\mathbb{R}^3$ along with the plane at infinity Π_∞. The usual Euclidean points, lines and planes in $\mathbb{R}^3$ along with the plane at infinity (and its points and lines) provide the geometric structure of $\mathbb{R}^3 \cup \Pi_\infty$ as a projective plane. This model is familiar to the Vision community through its central role in multiview geometry [32].

Via its correspondence with $\mathbb{P}^3$, rotation space $\mathrm{SO}(3)$ inherits the geometry of a projective space, wherein a "line" is the set of rotations corresponding 1-to-2 to a great circle in the quaternion sphere and a "plane" is the set of rotations corresponding 1-to-2 to a "great" 2-sphere in the (3-dimensional) quaternion sphere.

Many useful properties of $\mathrm{SO}(3)$ may be deduced using only the geometric properties of $\mathbb{P}^3$, and ignoring any of the algebraic properties (such as rotation multiplication), or the metric structure of $\mathrm{SO}(3)$, discussed in the next section. When considering the geometric properties of $\mathrm{SO}(3)$ in its embodiment as a projective space $\mathbb{P}^3$ we shall often find it convenient to refer to geometric concepts such as lines and planes, rather than circles and spheres in the quaternion sphere, or the corresponding curves and surfaces in $\mathrm{SO}(3)$. It will become apparent in the next section that these lines and planes in $\mathbb{P}^3$ in fact correspond to *geodesics* and *geodesic surfaces* in $\mathrm{SO}(3)$.

4 Distance Measures on SO(3)

We will be interested in distance measures (we use this term interchangeably for 'metric') on the group of rotations, which will give the rotations the structure of a metric space.

Bi-invariant distance. A distance measure $d : \mathrm{SO}(3) \times \mathrm{SO}(3) \to \mathbb{R}^+$ is called *bi-invariant* if

$$d(\mathtt{SR}_1, \mathtt{SR}_2) = d(\mathtt{R}_1, \mathtt{R}_2) = d(\mathtt{R}_1\mathtt{S}, \mathtt{R}_2\mathtt{S})$$

for all $\mathtt{S}$ and $\mathtt{R}_i$. Because of the homogeneous manifold structure of the rotation group (evidenced by the quaternion sphere), it is natural to be mostly interested in bi-invariant metrics. On $\mathrm{SO}(3)$, the following are the most common choices for the distance d.

Angular distance. Any rotation in $\mathrm{SO}(3)$ can be expressed as a rotation through a given angle θ about some axis. The angle can always be chosen such that $0 \leq \theta \leq \pi$, if necessary by reversing the direction of the axis. We define the angular distance between two rotations $\mathtt{R}$ and $\mathtt{S}$ to be the angle of the rotation $\mathtt{SR}^\top$, so chosen to lie in this range $[0, \pi]$. Thus,

$$d_\angle(\mathtt{S}, \mathtt{R}) = d_\angle(\mathtt{SR}^\top, \mathtt{I}) = \| \log(\mathtt{SR}^\top) \|_2$$

where the norm is the usual Euclidean norm in $\mathbb{R}^3$. Note that by this definition, the angular distance between two rotations is at most π. The angular distance function $d_\angle(\mathtt{S}, \mathtt{R})$ is equal to the rotation angle $\angle(\mathtt{SR}^\top)$. Note that we could equally well write $\mathtt{R}^\top\mathtt{S}$, $\mathtt{RS}^\top$ or $\mathtt{S}^\top\mathtt{R}$, since in all cases these represent a rotation through the same angle.

The angular distance between two rotations is easily computed from their quaternion representations. Thus, if $\mathbf{r}$ and $\mathbf{s}$ are quaternion representations of $\mathtt{R}$ and $\mathtt{S}$ respectively, and $\theta = d_\angle(\mathtt{S}, \mathtt{R})$, then

$$\theta = 2 \arccos(|c|) \text{ where } (c, \mathbf{v}) = \mathbf{s}^{-1} \cdot \mathbf{r} \ . \tag{5}$$

The absolute value sign in $|c|$ is required to account for the sign ambiguity in the quaternion representation of the rotation $\mathtt{S}^\top\mathtt{R}$. The positive sign is chosen so that the angle θ lies in the range $0 \leq \theta \leq \pi$, as required.

Once we have introduced the concept of geodesics in $\mathrm{SO}(3)$, we will also refer to angular distance as "geodesic distance," using these terms interchangeably.

Chordal distance. The *chordal distance* between two rotations $\mathtt{R}, \mathtt{S}$ in $\mathrm{SO}(3)$ is the Euclidean distance between them in the embedding space $\mathbb{R}^{3\times 3} = \mathbb{R}^9$. Thus,

$$d_{\text{chord}}(\mathtt{S}, \mathtt{R}) = \|\mathtt{S} - \mathtt{R}\|_{\text{F}}$$

where $\|\cdot\|_{\text{F}}$ represents the Frobenius norm of the matrix. This distance is easily related to the angular distance $\theta = d_\angle(\mathtt{S}, \mathtt{R})$ using Rodrigues' formula (3). Specifically, let $\mathtt{S}\mathtt{R}^\top = \exp(\theta\hat{\mathbf{v}})$. Since $[\hat{\mathbf{v}}]_\times$ and $[\hat{\mathbf{v}}]_\times^2$ are orthogonal to each other with respect to the Frobenius inner product, and since $\|[\hat{\mathbf{v}}]_\times\|_{\text{F}}^2 = \|[\hat{\mathbf{v}}]_\times^2\|_{\text{F}}^2 = 2$, formula (3) gives

$$\begin{aligned} d_{\text{chord}}(\mathtt{S}, \mathtt{R})^2 &= \|\mathtt{S} - \mathtt{R}\|_{\text{F}}^2 = \|\mathtt{S}\mathtt{R}^\top - \mathtt{I}\|_{\text{F}}^2 \\ &= 2(\sin^2(\theta) + (1 - \cos(\theta))^2) \\ &= 8\sin^2(\theta/2) \end{aligned}$$

from which we obtain the required relation

$$d_{\text{chord}}(\mathtt{S}, \mathtt{R}) = 2\sqrt{2}\,\sin(\theta/2)\ .$$

Quaternion distance. Another distance measure derives from the Euclidean distance between two quaternions in the embedding space $\mathbb{R}^4$. We may think to define a distance $d_{\text{quat}}(\mathtt{S}, \mathtt{R})$ between two rotations to be $d_{\text{quat}}(\mathtt{S}, \mathtt{R}) = \|\mathbf{s} - \mathbf{r}\|_2$, where $\mathbf{s}$ and $\mathbf{r}$ are quaternion representations of $\mathtt{S}$ and $\mathtt{R}$, respectively. Unfortunately, this simple equation will not do, since both $\mathbf{r}$ and $-\mathbf{r}$ represent the same rotation, and it is not clear which one to choose (and analogous for $\mathbf{s}$ and $-\mathbf{s}$, of course). However, this is resolved by defining

$$d_{\text{quat}}(\mathtt{S}, \mathtt{R}) = \min\{\|\mathbf{s} - \mathbf{r}\|_2, \|\mathbf{s} + \mathbf{r}\|_2\}$$

where the norm is the usual Euclidean norm in $\mathbb{R}^4$. Since quaternions satisfy the condition $\|\mathbf{s}\cdot\mathbf{t}\|_2 = \|\mathbf{s}\|_2\|\mathbf{t}\|_2$, where $\mathbf{s}\cdot\mathbf{t}$ represents the quaternion product, it is easily verified that the quaternion distance is bi-invariant.

The relationship of this to the angular distance is as follows. Let $\theta = d_\angle(\mathtt{S}, \mathtt{R}) = d_\angle(\mathtt{S}\mathtt{R}^\top, \mathtt{I})$ be the angle of the rotation $\mathtt{S}\mathtt{R}^\top$. Represent the identity rotation $\mathtt{I}$ by the quaternion $\mathbf{e} = (1, 0, 0, 0)$ and $\mathtt{S}\mathtt{R}^\top$ by the quaternion $\mathbf{s}\cdot\mathbf{r}^{-1} = (\cos(\theta/2), \hat{\mathbf{v}}\sin(\theta/2))$. Then the inner product of these two quaternions, considered simply as 4-vectors, is equal to $\cos(\theta/2)$. On the other hand, as an inner product of two unit vectors, it is equal to $\cos(\alpha)$, where α is the angle between the two vectors in $\mathbb{R}^4$. Thus, the angle between the two quaternions is $\alpha = \theta/2$. The distance $\|\mathbf{s}\cdot\mathbf{r}^{-1} - \mathbf{e}\|_2 = \|\mathbf{s} - \mathbf{r}\|_2$ is then equal to

$$d_{\text{quat}}(\mathtt{S}, \mathtt{R}) = 2\sin(\alpha/2) = 2\sin(\theta/4),$$

which is the distance between two unit vectors separated by an angle $\theta/2$. **Notation:** We will occasionally apply the angle metric $d_\angle(\cdot,\cdot)$ to quaternions, defining $d_\angle(\mathbf{s}, \mathbf{r}) = 2\alpha$ to be twice the angle between the two quaternions, considered as vectors in $\mathbb{R}^4$. Then for the corresponding rotations, $d_\angle(\mathtt{R}, \mathtt{S}) = \min(d_\angle(\mathbf{r}, \mathbf{s}), d_\angle(\mathbf{r}, -\mathbf{s}))$.

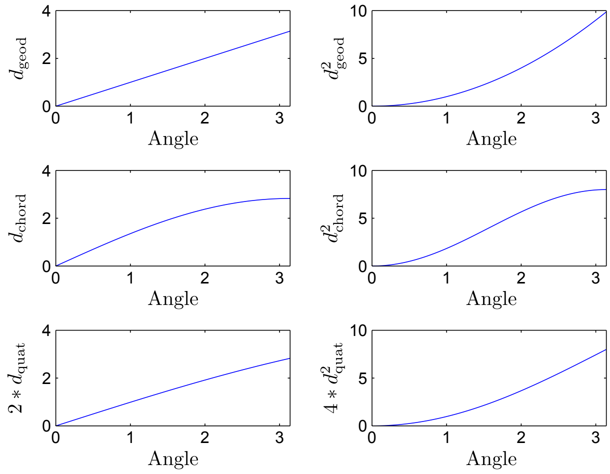


Fig. 2 Distance metrics. *On the left (top to bottom) are angular, chordal and quaternion distances plotted as a function of rotation angle. On the right the squared distances. Plots are shown for rotation angles from 0 to π. The plots of the quaternion metric are scaled to be comparable with the other metrics.*

Plots of the three different distance functions discussed so far, plotted as functions of the angular distance are shown in fig 2.

Distance in angle-axis space. Yet another distance on $\mathrm{SO}(3)$ may be defined as the Euclidean distance between corresponding vectors $\log(\mathtt{S})$ and $\log(\mathtt{R})$ in angle-axis space. However, if $\log(\mathtt{R})$ is taken to be the smallest length vector representing $\mathtt{R}$, then this metric is not continuous, in the sense that rotations through angles near π about opposite axes are not close to each other in this metric (but they are in the angle metric).

This problem can be resolved by choosing between alternative "branches" of the logarithm function. The definition then becomes

$$d_{\log}(\mathtt{S}, \mathtt{R}) = \min\|\mathbf{v}_r - \mathbf{v}_s\|_2$$

where the minimum is taken over all choices of vectors $\mathbf{v}_r$ and $\mathbf{v}_s$ such that $\exp[\mathbf{v}_r]_\times = \mathtt{R}$ and $\exp[\mathbf{v}_s]_\times = \mathtt{S}$.

With this definition, it can be shown [28] that

$$d_\angle(\mathtt{R}, \mathtt{S}) \leq d_{\log}(\mathtt{R}, \mathtt{S}) \leq (\pi/2)d_\angle(\mathtt{R}, \mathtt{S})\ ,$$

so both $d_\angle$ and $d_{\log}$ induce the same topology. However, the problem with this distance is that it is not bi-invariant, since $d_{\log}(\mathtt{TS}, \mathtt{TR}) \neq d_{\log}(\mathtt{S}, \mathtt{R})$ in general. We will have little occasion to use this metric.

4.1 Curve Length and Geodesics

We now, consider the meaning of curve length in a metric space, (M, d), where M is a set and d is the metric. We wish to do this for arbitrary curves, without any assumption of differentiability. A *curve* in M is a continuous function $\gamma : [0,1] \to M$; it joins the starting point $\gamma(0)$ to the end point $\gamma(1)$. The length of such a curve is defined as follows.

A *partition* of the interval $[0,1]$ is a sequence of points $0 = t_0 < t_1 < \ldots < t_{n-1} < t_n = 1$ in the interval $[0,1]$. We define

$$L(\gamma; \{t_i\}) = \sum_{i=1}^{n} d(\gamma(t_i), \gamma(t_{i-1})) \ .$$

It follows from the triangle inequality that if we refine the sequence $t_0, \ldots, t_n$ by adding extra points, then the value of $L(\gamma; \{t_i\})$ can not decrease.

Now, we define the length of the curve to be the supremum of $L(\gamma; \{t_i\})$ over all partitions. A curve for which this supremum is finite is called a *rectifiable* curve. Otherwise, the curve is considered to have infinite length.

Given two points $x, y \in M$, a *path* γ from x to y is a curve with $\gamma(0) = x$ and $\gamma(1) = y$. We may define a new metric on the space, called the *intrinsic metric* in which $\widehat{d}(x,y)$ is defined to be the infimum of the lengths of all paths from x to y. It is easily verified that this defines a metric on the space, and $\widehat{d}(x,y) \geq d(x,y)$.

We wish to find the relationship between the intrinsic metrics induced by two different metrics on the same space. The following result gives an answer.

Theorem 1 *If $d_1(x,y)$ and $d_2(x,y)$ are two metrics defined on a space M such that*

$$\lim_{d_1(x,y)\to 0} \frac{d_2(x,y)}{d_1(x,y)} = 1 \tag{6}$$

uniformly (with respect to x and y), then the length of any given curve is the same under both metrics. Consequently, the intrinsic metrics induced by d_1 and d_2 are identical.

The condition (6) is to be interpreted to mean that for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$1 - \varepsilon < \frac{d_2(x,y)}{d_1(x,y)} < 1 + \varepsilon \tag{7}$$

whenever x and y are chosen so that $d_1(x,y) < \delta$.

Now, consider a curve γ with length L_1 under the metric d_1, and L_2 under metric d_2; suppose both L_1 and L_2 are finite. Choose a value $\eta > 0$ and define $\varepsilon = \eta / L_1$. Let δ be chosen such that condition (7) is true. In this case

$$(1-\varepsilon) d_1(x,y) \leq d_2(x,y) \leq (1+\varepsilon) d_1(x,y)$$

provided $d_1(x,y) < \delta$.

Choose a partition $t_0, \ldots, t_n$ to satisfy

$$L_1 - \eta \leq \sum_{i=1}^{n} d_1(\gamma(t_{i-1}), \gamma(t_i)) \leq L_1 \ , \tag{8}$$

and

$$L_2 - \eta \leq \sum_{i=1}^{n} d_2(\gamma(t_{i-1}), \gamma(t_i)) \leq L_2 \ , \tag{9}$$

This can be achieved while at the same time making the partition sufficiently fine such that $d_1(\gamma(t_{i-1}), \gamma(t_i)) < \delta$ for all i. Then we have

$$\begin{aligned} L_2 &\geq \sum_{i=1}^{n} d_2(\gamma(t_{i-1}), \gamma(t_i)) \\ &\geq (1-\varepsilon) \sum_{i=1}^{n} d_1(\gamma(t_{i-1}), \gamma(t_i)) \\ &\geq (1-\varepsilon)(L_1 - \eta) \geq L_1 - 2\eta \end{aligned} \tag{10}$$

and

$$\begin{aligned} L_2 - \eta &\leq \sum_{i=1}^{n} d_2(\gamma(t_{i-1}), \gamma(t_i)) \\ &\leq (1+\varepsilon) \sum_{i=1}^{n} d_1(\gamma(t_{i-1}), \gamma(t_i)) \\ &\leq (1+\varepsilon) L_1 = L_1 + \eta \ . \end{aligned} \tag{11}$$

From (9), (10) and (11) it follows that $L_1 - 2\eta \leq L_2 \leq L_1 + 2\eta$. Since η was chosen arbitrarily, it follows that $L_1 = L_2$.

A slightly modified proof will be sufficient to show that if either L_1 or L_2 is infinite, then so is the other.

Equality of curve lengths. In the following exposition, we will use a convention that $\mathbf{r}$ and $\mathbf{s}$ represent unit quaternions, and that $\mathtt{R}$ and $\mathtt{S}$ are the corresponding rotation matrices.

The three metrics $d_\angle$, d_{quat} and d_{chord} defined on $\mathrm{SO}(3)$ are distinct, as we have shown. However, we wish to show that their induced intrinsic metrics are identical, up to scale. Let $\widehat{d}$ represent the intrinsic metric induced by a metric d. Letting $d_\angle(\mathtt{R}, \mathtt{S}) = \theta$, it was shown that $d_{\text{chord}}(\mathtt{R}, \mathtt{S}) = 2\sqrt{2}\, \sin(\theta/2)$. Therefore, it follows that

$$\lim_{d_\angle(\mathtt{R},\mathtt{S})\to 0} \frac{d_{\text{chord}}(\mathtt{R}, \mathtt{S})}{\sqrt{2}\, d_\angle(\mathtt{R}, \mathtt{S})} = 1 \ .$$

From this, theorem 1 implies that for a given curve in $\mathrm{SO}(3)$, the curve lengths measured with respect to the angle and chordal metrics differ by a constant factor $\sqrt{2}$. Since the induced intrinsic metrics are defined as the infimum of path lengths, it follows that $\widehat{d}_{\text{chord}}(\mathtt{R}, \mathtt{S}) = \sqrt{2}\, \widehat{d}_\angle(\mathtt{R}, \mathtt{S})$.

Similarly, we know that $d_{\text{quat}}(\mathtt{R}, \mathtt{S}) = 2\sin(\theta/4)$, and so by the same argument, we see that $\widehat{d}_{\text{quat}}(\mathtt{R}, \mathtt{S}) = (1/2)\widehat{d}_\angle(\mathtt{R}, \mathtt{S})$. We have shown the following result.

Theorem 2 *Let $\gamma(t)$ be a curve in $\mathrm{SO}(3)$ and define $L_{\text{quat}}(\gamma)$, $L_{\text{chord}}(\gamma)$ and $L_\angle(\gamma)$ to be the curve lengths with respect to the three different metrics. Then*

$$L_{\text{chord}}(\gamma) = 2\sqrt{2}\, L_{\text{quat}}(\gamma) = \sqrt{2}\, L_\angle(\gamma) \ .$$

For the induced intrinsic metrics,

$$\widehat{d}_{\text{chord}}(\mathtt{R}, \mathtt{S}) = 2\sqrt{2}\, \widehat{d}_{\text{quat}}(\mathtt{R}, \mathtt{S}) = \sqrt{2}\, \widehat{d}_\angle(\mathtt{R}, \mathtt{S}) \ .$$

The quaternion metric on $\mathrm{SO}(3)$ is derived from the Euclidean distance metric on the quaternion sphere. In fact, the two metrics are locally equal. It follows that the length of a curve on S^3 under the Euclidean metric is the same as the length L_{quat} of the corresponding curve on $\mathrm{SO}(3)$.

Note that the angle metric $d_\angle$ is identical with its induced intrinsic metric $\widehat{d_\angle}$. In standard terminology, this is expressed by saying that $(\mathrm{SO}(3), d_\angle)$ is a *length metric space*. This is not true for the other metrics d_{chord} and d_{quat}.

Geodesics. A geodesic is defined to be a locally length-minimizing path. To be more specific, let I be an interval in $\mathbb{R}$; a path is a continuous function $\gamma : I \rightarrow M$ for any metric space (M, d). We allow I to be infinite at either end, to allow infinite paths. The path γ is a geodesic if there exist open intervals I_i covering I such that for any two points x and y in I_i, the path γ restricted to the interval $[x, y]$ is a shortest path from $\gamma(x)$ to $\gamma(y)$.

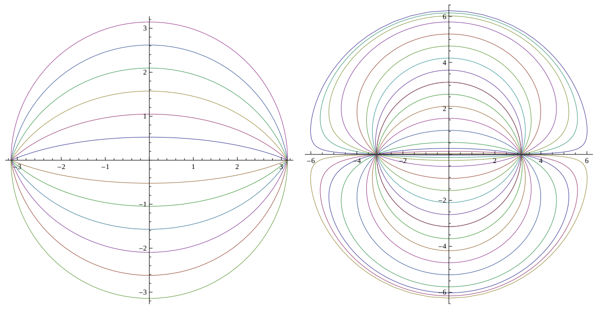
It is well known that the shortest path between two points on the 3-sphere S^3 lies on a great circle. Moreover, two points on S^3 may be joined by a path that achieves the shortest length. Since path lengths in $\mathrm{SO}(3)$ are equal (up to a scale factor depending on the metric being used) to path lengths on the quaternion sphere, it follows that any two points in $\mathrm{SO}(3)$ may also be joined by a minimum length geodesic. This result, obvious enough in $\mathrm{SO}(3)$, is true under very general conditions, as expressed in the Hopf-Rinow theorem (see Theorem 7.1 in [61] for a very general version), which states that if a length-metric space (M, d) is complete and locally compact then any two points in M can be connected by a minimizing geodesic.

We now consider more explicitly the shape of $\mathrm{SO}(3)$-geodesics as they appear in our main representations of $\mathrm{SO}(3)$ as a group of rotation matrices, the quaternion sphere and angle-axis space.

4.1.1 Geodesics in the quaternion sphere.

As we observed above, the great circles on S^3 are the geodesics. For varying t, the curve $\gamma(t) = (\cos(t\theta/2), \sin(t\theta/2)\hat{\mathbf{v}})$ is the great circle in the quaternion sphere S^3 passing through the points $(1, \mathbf{0})$ and $\mathbf{s} = (\cos(\theta/2), \sin(\theta/2)\hat{\mathbf{v}})$. Multiplication by a quaternion $\mathbf{r}$ represents a rigid transformation of the quaternion sphere. Consequently the curve $\mathbf{r} \cdot \gamma(t)$ is also a great circle on S^3, passing through $\mathbf{r}$ and $\mathbf{r} \cdot \mathbf{s}$. This is the general form of a quaternion great circle; any geodesic in S^3 is of the form

$$\gamma(t) = \mathbf{r} \cdot (\cos(t\theta/2), \sin(t\theta/2)\hat{\mathbf{v}}).$$



(a) (b)

Fig. 3 *Geodesics in angle axis space: (a) geodesics lying in the ball B_π; (b) Geodesics extended in angle-axis space form closed curves. These curves correspond via a one-to-one mapping with the great circles on the quaternion sphere.*

4.1.2 Geodesics in angle-axis space.

The curve in angle-axis space corresponding to the geodesic $\gamma(t) = (\cos(t/2), \sin(t/2)\hat{\mathbf{v}})$ in the quaternion sphere is the curve given by $t\hat{\mathbf{v}}$, namely a straight line through the origin.

It is useful to understand what arbitrary geodesics in angle-axis space look like. (It should be understood that when we talk of geodesics in angle-axis space or another representation of rotations, we mean curves that correspond to geodesics in $\mathrm{SO}(3)$).

The shape of geodesics in angle-axis space is shown in fig 3 which shows sample geodesics lying in some plane in angle-axis space. Geodesics through the origin (identity rotation) will be radial lines in angle-axis space. Other geodesics will be curves (neither circles nor ellipses) passing through any pair of opposite points on the boundary of B_π, both these points representing the same rotation.

It is interesting to see (fig 3) that geodesics can be extended beyond the ball B_π, representing rotations through angles greater than π. As the figure shows the geodesics will close to form closed curves in angle-axis space.

4.1.3 Geodesics in $\mathrm{SO}(3) \subset \mathrm{GL}(3)$

Mapping the geodesic $\mathbf{r} \cdot (\cos(t/2), \sin(t/2)\hat{\mathbf{v}})$ we obtain the geodesic in $\mathrm{SO}(3)$, namely $\mathtt{R}\exp[t\hat{\mathbf{v}}]_\times$.

The shortest path in $\mathrm{SO}(3)$ from rotation $\mathtt{R}$ to $\mathtt{S}$ is given by

$$\gamma(t) = \mathtt{R}\exp(t\log(\mathtt{R}^\top\mathtt{S})), \tag{12}$$

which is a one-parameter family of rotations about a single axis.

4.1.4 Geodesics and the Gnomonic Projection

The gnomomic projection, described in section 3.4 has the particularly pleasing property that it maps geodesics in $\mathrm{SO}(3)$ to geodesics (straight lines) in $\mathbb{R}^3$. As noted, an

$$
\begin{aligned}
d_\angle(\mathtt{S},\mathtt{R}) &= \theta \\
d_{\text{chord}}(\mathtt{S},\mathtt{R}) &= 2\sqrt{2}\,\sin(\theta/2) \\
d_{\text{quat}}(\mathtt{S},\mathtt{R}) &= 2\sin(\theta/4) \\
d^2_\angle(\mathtt{S},\mathtt{R}) &= \theta^2 \\
d^2_{\text{chord}}(\mathtt{S},\mathtt{R}) &= 8\sin^2(\theta/2) = 4(1-\cos(\theta)) \\
d^2_{\text{quat}}(\mathtt{S},\mathtt{R}) &= 4\sin^2(\theta/4) = 2(1-\cos(\theta/2)) \\
\widehat{d}_\angle(\mathtt{S},\mathtt{R}) &= \theta \\
\widehat{d}_{\text{chord}}(\mathtt{S},\mathtt{R}) &= \sqrt{2}\,\theta \\
\widehat{d}_{\text{quat}}(\mathtt{S},\mathtt{R}) &= \theta/2
\end{aligned}
$$

Table 2 *Relationship between the different metrics on* $\mathrm{SO}(3)$.

$\mathrm{SO}(3)$-geodesic, when represented in the quaternion sphere is just a great circle. Such a great circle is formed by the intersection of a 2-dimensional plane (linear space) in $\mathbb{R}^4$ with the unit sphere. Therefore, the projection of a great circle from the centre of the sphere is just a 2-dimensional plane. When intersected with the tangent plane at some point on the sphere (the centre of the gnomonic projection), it forms a straight line.

This correspondence of geodesics with straight lines in $\mathbb{R}^3$ allows us to reason about geodesics in $\mathrm{SO}(3)$, and also gives us a simple intuitive understanding of these geodesics.

4.1.5 Summary

We will chiefly be interested in three distance functions and their squares. These are as follows.

1. Angular distance $d_\angle(\mathtt{S},\mathtt{R})$, equal to the angle θ belonging to the rotation $\mathtt{S}\mathtt{R}^\top$. When equipped with this metric, $\mathrm{SO}(3)$ is a length metric space. This seems to be the most natural metric for $\mathrm{SO}(3)$.
2. Chordal distance $d_{\text{chord}}(\mathtt{S},\mathtt{R})$, the distance inherited from the embedding of the rotations in $\mathbb{R}^{3\times 3} = \mathbb{R}^9$, equipped with the usual Euclidean metric.
3. Quaternion distance $d_{\text{quat}}(\mathtt{S},\mathtt{R})$ induced by the identification of rotations with points on the unit quaternion sphere, with metric inherited from $\mathbb{R}^4$.

The intrinsic metrics induced by these three metrics are, apart from constant scale factors, all the same and equal to the angle metric. The scale differences between the three intrinsic metrics are a source of potential confusion and irritation. Table 2 gives the values of the different metrics in terms of the angular distance.

The different induced intrinsic metrics $\widehat{d}$ determine the length of paths in rotation space, including the length of geodesics. Because of the differences in scale the length of paths is ambiguous. To settle this, we choose the angle metric as being the standard and most natural metric

– When we talk of length of paths or distances in rotation space, we mean path length or distance under the angle metric.

In addition, we will frequently refer to angular distance between two rotations as the *geodesic distance*, the length along the shortest geodesic path from one to the other.

4.2 The Cosine Rule in $\mathrm{SO}(3)$.

In planar geometry, the cosine rule states that $c^2 = a^2 + b^2 - 2ab\cos(C)$, where a, b, and c are the sides of a triangle and C is the angle opposite c. We wish to have a similar formula for geodesic triangles in $\mathrm{SO}(3)$.

Proposition 2 *Let a, b and c be the lengths of three geodesic line segments in $\mathrm{SO}(3)$, forming a triangle with vertices* $\mathtt{A}$, $\mathtt{B}$ *and* $\mathtt{C}$*. If c is the length of the smaller geodesic arc between* $\mathtt{A}$ *and* $\mathtt{B}$*, then*

$$
\cos(\frac{c}{2}) = \left|\cos(\frac{a}{2})\cos(\frac{b}{2}) + \sin(\frac{a}{2})\sin(\frac{b}{2})\cos(\widehat{\mathtt{C}})\right| , \quad (13)
$$

where $\widehat{\mathtt{C}}$ *is the angle at vertex* $\mathtt{C}$*.*

Notes. This rule is true whether a and b are the shorter geodesic lengths or not, as long as the angle $\widehat{\mathtt{C}}$ is measured between the two corresponding geodesic arcs meeting at $\mathtt{C}$. Note also that the length of the longer geodesic arc c' between $\mathtt{A}$ and $\mathtt{B}$ satisfies $\cos(c'/2) = -\cos(c/2)$.

We assume without loss of generality that $\mathtt{C}$ is the identity rotation and consider the representation of the rotations in angle-axis space, with $a\,\hat{\mathbf{v}}$ and $b\,\hat{\mathbf{w}}$ being the representations of $\mathtt{A}$ and $\mathtt{B}$ respectively. Here $\hat{\mathbf{v}}$ and $\hat{\mathbf{w}}$ are unit vectors. The geodesic arcs from $\mathtt{C}$ to $\mathtt{A}$ and $\mathtt{B}$ correspond to the radial line segments from the origin to $a\,\hat{\mathbf{v}}$ and $b\,\hat{\mathbf{w}}$ respectively, and $\widehat{\mathtt{C}}$ is simply the angle between these line segments at the origin, so $\cos(\widehat{\mathtt{C}}) = \langle\hat{\mathbf{v}},\, \hat{\mathbf{w}}\rangle$.

The required value c is simply the angular distance between rotations $\mathtt{A}$ and $\mathtt{B}$. This may be computed using quaternion multiplication. Let $\mathbf{a} = (\cos(a/2), \hat{\mathbf{v}}\sin(a/2))$, and $\mathbf{b} = (\cos(b/2), \hat{\mathbf{w}}\sin(b/2))$ be the quaternion representations of $\mathtt{A}$ and $\mathtt{B}$. Calculating in quaternions

$$
\mathbf{a}\cdot\mathbf{b}^{-1} = (\cos(a/2)\cos(b/2) + \sin(a/2)\sin(b/2)\,\langle\hat{\mathbf{v}},\, \hat{\mathbf{w}}\rangle\,, \ldots)
$$

where we do not need to compute the second part of the quaternion. The required formula for $\cos(c/2)$ now follows directly from (5).

5 Single Rotation Averaging

We now have the machinery to be able to consider each of the rotation averaging problems described in section 1. First, we consider (single) rotation averaging in $\mathrm{SO}(3)$ under the various different metrics of interest. Given n rotations $\mathtt{R}_i$, the problem is to find the rotation $\mathtt{R}$ that minimizes the cost function

$$C(\mathtt{R}) = \sum_{i=1}^{n} d(\mathtt{R}_i, \mathtt{R})^p \ , \tag{14}$$

where d is one of our metrics, and $p = 1$ or 2.

In this section, we will analyze the convexity of $C(\mathtt{R})$ on $\mathrm{SO}(3)$ and give convergent algorithms for the various metrics. The reader unfamiliar with the concept of (geodesic) convexity on Riemannian manifolds is referred to the appendix. It also contains a detailed discussion of weak convexity, an essential generalization of convexity that allows for larger domains of convexity.

5.1 The Geodesic and Quaternion Means

For the geodesic (angle) metric, the associated L_2-mean is usually called the Karcher mean [26] or the geometric mean [57]. A necessary condition [57, (3.12)] for $\mathtt{R}$ to be a $d_\angle^2$-mean of $\{\mathtt{R}_1, \ldots, \mathtt{R}_n\}$ is given by

$$\frac{1}{n}\sum_{i=1}^{n} \log(\mathtt{R}^\top \mathtt{R}_i) = 0. \tag{15}$$

For the L_2 geodesic or quaternion metrics, an individual term $d^2(\mathtt{R}, \mathtt{R}_i)$ in (14) is strictly convex as a function of $\mathtt{R}$ on an open ball $\mathring{B}(\mathtt{R}_i, \pi)$, and hence the cost function $C(\mathtt{R})$ is strictly convex on any connected component of the intersection of the open balls $\mathring{B}(\mathtt{R}_i, \pi)$. In general, the intersection of open balls consists of several components, as shown in proposition 11 in the appendix. A given open ball $\mathring{B}(\mathtt{R}_i, \pi)$ consists of the whole of $\mathrm{SO}(3)$, except for the plane consisting of rotations at angular distance π from $\mathtt{R}_i$. It follows that the total cost function (14) is strictly convex except on the union of these planes. This, and a little more is stated in the following theorem.

Theorem 3 *Let $d(\cdot,\cdot)$ be the geodesic or quaternion metric on $\mathrm{SO}(3)$. Given rotations $\mathtt{R}_i$, $i = 1, \ldots, n$, the cost function $C(\mathtt{R}) = \sum_{i=1}^n d(\mathtt{R}_i, \mathtt{R})^2$ is strictly convex, except on the union of planes*

$$\Pi_i = \{\mathtt{S} \in \mathrm{SO}(3) \,|\, d_\angle(\mathtt{R}_i, \mathtt{S}) = \pi\}$$

in the following sense. These sets Π_i divide $\mathrm{SO}(3)$ into at most $\binom{n}{3} + n$ regions whose interior is weakly convex. $C(\mathtt{R})$ is strictly convex on the interior of each of these regions and is non-differentiable on the boundary, that is, on the union of the sets Π_i. The cost function $C(\mathtt{R})$ has at most one minimum on each of the regions and hence there are at most $\binom{n}{3} + n$ minima.

Proof. Each individual cost function $d(\mathtt{R}_i, \mathtt{R})^2$ is strictly convex on the open ball $\mathring{B}(\mathtt{R}_i, \pi)$. Since each $\mathring{B}_i(\mathtt{R}_i, \pi)$ is weakly convex, their intersection consists of at most $\binom{n}{3} + n$ weakly convex components by proposition 11. Each $d(\mathtt{R}_i, \mathtt{R})^2$ is strictly convex on each such component. Hence their sum is strictly convex, and has a unique minimum on each component (by proposition 20). The proof will be completed by showing that $C(\mathtt{R})$ cannot have a minimum on the set $\mathrm{SO}(3) \setminus \bigcap_{i=1}^n \mathring{B}(\mathtt{R}_i, \pi) = \bigcup_{i=1}^n \Pi_i$.

Consider a point $\mathtt{S}$ in $\bigcup_{i=1}^n \Pi_i$ and choose a geodesic through $\mathtt{S}$ that does not lie on any of the planes Π_i. For those i such that $\mathtt{S} \in \Pi_i$, the function $d(\mathtt{R}_i, \mathtt{S})^2$ restricted to the geodesic has an upward cusp at $\mathtt{S}$, whereas for those i such that $\mathtt{S} \notin \Pi_i$ the same function is smooth. The sum of two such functions cannot be a minimum, so $\mathtt{S}$ is not a minimum of $C(\mathtt{R})$. This completes the proof. □

Theorem 3 indicates that $\mathrm{SO}(3)$ may be divided into a large number of individual weakly convex regions, each with its own local minimum. It may seem, therefore, that the problem of finding the global minimum is quite challenging. The following observation shows that if the rotations $\mathtt{R}_i$ are not too widely separated, one of the weakly convex regions may be quite large. The following result follows directly from theorem 3 and proposition 19 in the appendix. See there for the notion of convex basin $B^\natural$.

Theorem 4 *Let $d(\cdot,\cdot)$ be the geodesic or quaternion metric on $\mathrm{SO}(3)$. Given rotations $\mathtt{R}_i$, $i = 1, \ldots, n$, all lying in a weakly convex set B, the cost function $C(\mathtt{R}) = \sum_{i=1}^n d(\mathtt{R}_i, \mathtt{R})^2$ is strictly convex on the convex basin $B^\natural$, and hence has at most a single isolated minimum on $B^\natural$.*

The most important case is when B is convex, in which case $B^\natural$ is a weakly convex set containing B (proposition 19). If B is an open ball $\mathring{B}(\mathtt{S}, r)$ with $r \leq \pi$, then $B^\natural = B(\mathtt{S}, \pi - r)$, so if r is small, then the cost function is strictly convex on a large ball. The special case of the geodesic metric and $r = \pi/2$ is classical, see Theorem 3.7 in [26], and we restate it in the following corollary.

Corollary 1 *Let $\mathtt{R}_i$ be rotations satisfying $d_\angle(\mathtt{R}_i, \mathtt{S}) < \pi/2$ for some rotation $\mathtt{S}$ and for all i, then*

$$C(\mathtt{R}) = \sum_{i=1}^{n} d_\angle(\mathtt{R}_i, \mathtt{R})^2$$

is strictly convex on $B(\mathtt{S}, \pi/2)$, and hence has a single isolated minimum on that set.

Note that in general we do not claim that the cost function does in fact have even a local minimum on $B^\natural$. In fact it is not difficult to find an example where there is no such minimum, in the case where B is weakly convex, but not convex. It will be shown in the next section however, that if B is convex, then a unique local minimum, in fact the global minimum of the cost function lies in B.

5.2 The Global Minimum

In the previous section, we identified the regions on which the geodesic or quaternion L_2 cost functions are strictly convex, and indicated the existence of multiple possible local minima. According to theorem 4, if the rotations all lie in a convex set B, then the cost function is strictly convex on $B^\natural$, which is a weakly convex set containing B (according to proposition 19). In the next theorem we give a much stronger result, showing that in fact the *global minimum* of the cost function lies in B. In fact this will be shown in a more general framework that applies to all the metrics that we are considering in this paper, and more.

Theorem 5 *Let B be a convex subset of $\mathrm{SO}(3)$ and let the rotations $\mathtt{R}_i$, $i = 1, \ldots, n$ be contained in B. Let $d_i(\mathtt{R})$ be any strictly monotonic function of (geodesic) distance $d_\angle(\mathtt{R}_i, \mathtt{R})$. Then any global minimum in $\mathrm{SO}(3)$ of the function*

$$C_f(\mathtt{R}) = \sum_{i=1}^{n} d_i(\mathtt{R})$$

lies in B.

By strictly monotonic here, we mean that $d_i(\mathtt{R}) < d_i(\mathtt{R}')$ if and only if $d_\angle(\mathtt{R}_i, \mathtt{R}) < d_\angle(\mathtt{R}_i, \mathtt{R}')$. Examples include any of the L_p distance metrics we consider in this paper (including those listed in theorem 11), also weighted distances $d_i(\mathtt{R}) = w_i\, d_\angle^p(\mathtt{R}_i, \mathtt{R})$ for weights $w_i > 0$, as well as robust functions such as Huber distance and others ([32]). In all these cases, the theorem shows that if rotations $\mathtt{R}_i$ all lie in a convex set, then their "mean" under any of these "generalized distance" functions also lies in the convex set.

Furthermore, it was shown in theorem 4 that if $d_i(\mathtt{R})$ is the L_2 geodesic or quaternion metric there exists a single local, and hence by this theorem a unique global minimum in B. For the other metrics listed in theorem 11, the present theorem holds, but as will be seen later, there is not necessarily a *unique* global minimum in B.

If our intention were to prove this theorem in $\mathbb{R}^n$, then the result would be intuitively obvious and the proof simple. One could argue as follows. If $\mathbf{X}$ is a point not lying in a closed convex set B', then there exists a plane Π separating $\mathbf{X}$ from B'. Let $\mathbf{N}$ be a normal vector to the plane, pointing from $\mathbf{X}$ perpendicular and towards the plane Π. Then the distance from $\mathbf{X}$ to any point $\mathbf{Y}_i$ in B' decreases in the direction $\mathbf{N}$. Therefore, $\mathbf{X}$ can not be a minimum of $\sum_{i=1}^{n} d_i(\mathbf{X})$, for any increasing function $d_i(\mathbf{X})$ of the distance from $\mathbf{X}$ to $\mathbf{Y}_i$. Since the convex hull of the rotations $\mathtt{R}_i$ is a closed convex set, this argument shows that the minimum must lie in this convex hull, and hence in any convex set B containing all the $\mathtt{R}_i$. This proof does not work in $\mathrm{SO}(3)$, since the distance of $\mathbf{X}$ to points in B' does not necessarily decrease in the direction $\mathbf{N}$.

Neither is the theorem true for rotations in a weakly convex set. It is easy to find counterexamples. For instance, consider the closed ball of radius $5\pi/6$ about the identity rotation and let $\mathtt{R}_1$ and $\mathtt{R}_2$ be rotations through $5\pi/6$ and $-5\pi/6$ about some axis, both lying in this ball. However, the rotation $\mathtt{R}$, through angle π about this axis is the L_2-mean, minimizing the sum of squared distances to $\mathtt{R}_1$ and $\mathtt{R}_2$, since $d_\angle(\mathtt{R}_i, \mathtt{R}) = \pi/6$ for $i = 1, 2$.

It is remarkable that there are counterexamples to this theorem for manifolds other than $\mathrm{SO}(3)$, see [11], although it has been shown to hold for the special case of a set of points that are contained within a small ball. More specifically, Le [47] studied geodesic L_2 averaging on general Riemannian manifolds and showed the existence of a unique global L_2-mean of a set of points contained in an open ball of radius at most $\pi/4$ (this is the numerical value on $\mathrm{SO}(3)$ of the general bound given in Le's paper). This result was improved by Afsari [2] who achieved a radius bound of $\pi/2$ (on $\mathrm{SO}(3)$) and derived analogous results for general L_p-means. Afsari also studied convex sets but only those contained in a small ball. Nevertheless, the theorem is true for all convex sets in $\mathrm{SO}(3)$.

Theorem 5 will be proved as an easy consequence of the following two lemmas. Proofs are provided in the appendix.

Lemma 1 *Theorem 5 is true in the special case where B is a closed convex set and the rotations $\mathtt{R}_i$ lie in the interior of B.*

Lemma 2 (Pumping lemma.) *Let B be a closed convex subset of $\mathrm{SO}(3)$ then there exists a larger closed convex subset $\hat{B}$ of $\mathrm{SO}(3)$ such that all points of B lie in the interior of $\hat{B}$. Furthermore, the intersection of all such sets $\hat{B}$ is equal to B.*

In a sense, we may pump up B, like a balloon to form a larger closed convex set. (We recognize that the term "pumping lemma" is used in the literature for an entirely different result, but there should be no confusion.)

The theorem follows directly from these two results. Indeed, if rotations $\mathtt{R}_i; i = 1, \ldots, n$ lie in a convex set B, then their convex hull $H \subset B$ exists and is closed. In this case, the pumping lemma shows that there exists a closed convex set $\hat{B}$ containing the points $\mathtt{R}_i$ in its interior. Then lemma 1 will hold for $\hat{B}$, guaranteeing that the mean of the $\mathtt{R}_i$ lies in $\hat{B}$ for all such $\hat{B}$ containing H. However, by the second part of the pumping lemma, the mean must lie in H, and hence in B.

5.3 The Geodesic L_2-mean

The rotation minimizing $C(\mathtt{R}) = \sum_{i=1}^{n} d_\angle(\mathtt{R}, \mathtt{R}_i)^2$ is also known at the Karcher mean of the rotations. Manton [53] has provided a convergent algorithm to find this mean, where the inner loop of the algorithm is computing the average in the tangent space, and then projecting back onto the manifold $\mathrm{SO}(3)$ via the exponential map. Note that Condition (15) is a necessary condition for $\mathtt{R}$ to minimize this cost function. The algorithm is as follows.

```
Set R := R_1. Choose a tolerance ε > 0.
loop
    Compute r := (1/n) Σ_{i=1}^n log(R^⊤ R_i).
    if ||r|| < ε then
        return R
    end if
    Update R := R exp(r).
end loop
```

Algorithm 1: *Computing the geodesic L_2-mean on* $\mathrm{SO}(3)$

In fact, this algorithm is shown to be an instance of simple Riemannian gradient descent (with constant step-size 1) and it is shown that if all the rotations lie in a closed ball of radius $\delta < \pi/2$, then an implementation with arbitrary numerical accuracy would terminate within a $d_\angle$-distance of $\varepsilon \tan(\delta)/\delta$ of the mean [53, Theorem 5]. See also [48] for similar results.

Our convexity results imply that if the rotations lie in an arbitrary convex set B, then a gradient descent algorithm with properly chosen step-size will converge to the global minimum [62].

Higher order algorithms. Second and other higher order algorithms for means on manifolds appear to be much less studied than first order algorithms like gradient descent. For a Newton-type algorithm to compute the Karcher mean see [45]. While a Riemannian generalization of the popular BFGS method is well known and has been stated specifically for compact Stiefel manifolds in [67], it appears not to have been applied to the special case of rotation averaging. The same holds for conjugate gradient [16].

There does not appear to be any non-iterative algorithm to solve the geodesic L_2 single rotation averaging problem.

5.4 The Geodesic L_1-mean

Another interesting mean with respect to the angular distance $d_\angle$ is the associated L_1-mean, that is, the global minimum of the function

$$C(\mathtt{R}) = \sum_{i=1}^{n} d_\angle(\mathtt{R}_i, \mathtt{R}). \qquad (16)$$

We might assume the L_1-mean to be more robust to errors than the corresponding L_2-mean. See [12] for some evidence for this assertion.

However, this minimizer is not always unique. Take for example any geodesic $\gamma : I \to \mathrm{SO}(3)$ of length less than $\pi/2$ and take $\mathtt{R}_1 = \gamma(t_1)$ and $\mathtt{R}_2 = \gamma(t_2)$, where $t_1, t_2 \in I$. Then any point $\gamma(t)$, $t \in [t_1, t_2]$ on the geodesic yields the same minimal cost $C(\gamma(t)) = C(\gamma(t_1))$. Note further that $C(\mathtt{R})$ is not differentiable at the points $\mathtt{R} = \mathtt{R}_i, i = 1, \ldots, n$. Hence not all of the minimizers are critical points of $C(\mathtt{R})$ in this example.

While theorem 11 merely states that the angular distance is convex on open balls of radius π, a more careful evaluation of the Hessian (see Table 3) and its eigendirections reveals that in fact the angular distance is strictly convex along any geodesic, except for the geodesics that pass through the reference point $\mathtt{R}_i$. Thus, unless all the rotations $\mathtt{R}_i$ lie on a single geodesic, the cost function will be strictly convex along any geodesic. This means that all the theorems from the previous section apply under this additional condition, and we get the following strong result.

Theorem 6 *Let $\mathtt{R}_i$ be rotations not all lying on a single geodesic. Let $S = \{\mathtt{R}_1, \ldots, \mathtt{R}_n\}$. Then, the cost function $C(\mathtt{R}) = \sum_{i=1}^n d_\angle(\mathtt{R}_i, \mathtt{R})$ is differentiable everywhere in $S^\natural$ except at the points $\mathtt{R}_i$. It is strictly convex everywhere on $S^\natural$ and has at most one local minimum in the closure of each of the $\binom{n}{3} + n$ connected components of $S^\natural$.*

If all $\mathtt{R}_i$ lie in a convex set B then $C(\mathtt{R})$ is strictly convex on the weakly convex set $B^\natural$ containing B, and the unique global geodesic L_1-mean lies in B.

A practical algorithm. We propose a Riemannian gradient descent algorithm with geodesic line search to compute the L_1-mean. A detailed derivation of the gradient

$$\nabla C(\mathtt{R}) = -\mathtt{R} \sum_{i=1}^{n} \frac{\log(\mathtt{R}^\top \mathtt{R}_i)}{\| \log(\mathtt{R}^\top \mathtt{R}_i) \|} \qquad (17)$$

is given in the appendix.

The algorithm starts at some initial point $\mathtt{R}$; how this is chosen is discussed below. It then iteratively makes steps in the direction of the downhill gradient, using line search to find the minimum in the gradient direction, and continuing until convergence.

The search direction is computed using (17). Note that this formula is invalid if $\mathtt{R}$ is equal to one of the $\mathtt{R}_i$ – the cost function is not differentiable at this point. The following observation allows us to compute the downhill gradient direction in this case.

Let $\mathtt{R} = \mathtt{R}_i$, equal to one of the rotations being averaged, and define

$$\mathbf{r} = \sum_{\mathtt{R}_i \neq \mathtt{R}} \log(\mathtt{R}^\top \mathtt{R}_i) / \| \log(\mathtt{R}^\top \mathtt{R}_i) \|, \qquad (18)$$

namely the gradient formula (17), omitting the term involving $\mathtt{R}_i$. If $\|r\| \leq 1$, then $\mathtt{R}$ is a local minimum of the cost function (hence the global minimum if all $\mathtt{R}_i$ lie in a convex set and not all on a single geodesic, according to theorem 6). Otherwise, $\mathbf{r}$ is a vector in the direction of most rapid decrease of (17), so we may use $\mathbf{r}$ so defined as the search direction.

This observation is easily verified, since the term involving $\mathtt{R}_i$ missing from (18) corresponds to the gradient of the function $d_\angle(\mathtt{R}, \mathtt{R}_i)$. The gradient of this function points everywhere radially away from $\mathtt{R}_i$, with magnitude 1.

If all the rotations $\mathtt{R}_i$ lie in a convex set B, we define an initial rotation $\mathtt{R}^{(0)} = \text{argmin}_{\{\mathtt{R}_i\}}\ C(\mathtt{R}_i)$ and set $\alpha^{(0)} = C(\mathtt{R}^{(0)})$. Then the sublevel set

$$S(\alpha^{(0)}, B) = \{\mathtt{R} \in B \mid C(\mathtt{R}) < \alpha^{(0)}\}$$

is a convex set on which the cost function is differentiable (since it does not contain any $\mathtt{R}_i$) and convex, and achieves its global minimum. If we start the iteration by setting $\mathtt{R}$ equal to $\mathtt{R}^{(0)}$ then either this is the required minimum or else the first step in the direction (18) will place us inside the sublevel set, and no future step will take us to (or pass through) any of the rotations. This will allow us to compute gradients without fear, and use gradient-based line search algorithms if desired. Convergence follows directly from [1, Corollary 4.3.2].

The complete algorithm is as follows.

1: Choose a tolerance $\varepsilon > 0$.
2: Set $\mathtt{R} := \text{argmin}_{\{\mathtt{R}_i\}}\ C(\mathtt{R}_i)$.
3: Compute $\mathbf{r} := \sum_{\mathtt{R}_i \neq \mathtt{R}} \log(\mathtt{R}^\top \mathtt{R}_i)/\|\log(\mathtt{R}^\top \mathtt{R}_i)\|$.
4: **if** $\|\mathbf{r}\| \leq 1$ **then**
5:     **return** $\mathtt{R}$
6: **else**
7:     **loop**
8:         Compute $s^* := \text{argmin}_{s \geq 0}\ C(\mathtt{R}\exp(s\mathbf{r}))$.
9:         **if** $\|s^*\mathbf{r}\| < \varepsilon$ **then**
10:             **return** $\mathtt{R}$
11:         **end if**
12:         Update $\mathtt{R} := \mathtt{R}\exp(s^*\mathbf{r})$.
13:         Compute $\mathbf{r} := \sum_{i=1}^n \log(\mathtt{R}^\top \mathtt{R}_i)/\|\log(\mathtt{R}^\top \mathtt{R}_i)\|$.
14:     **end loop**
15: **end if**

Algorithm 2: *Computing the geodesic L_1-mean of a set of rotations $\mathtt{R}_i$. If all rotations lie inside a convex set B, then this algorithm is guaranteed to converge.*

Possibly, the easiest way to implement the line search in Step 8 is a Fibonacci search on a large enough interval, though gradient-based search is also a possibility. We suggested initializing at the best rotation $\mathtt{R}_i$, but this may be expensive with many rotations, and is probably not necessary, as long as (18) is used to compute the search direction in the case where $\mathtt{R}$ is equal to one of the $\mathtt{R}_i$. An attractive alternative is to initialize the algorithm with the geodesic L_2-mean, which is within the convex basin containing the global minimum, if all rotations $\mathtt{R}_i$ lie in a convex set B.

5.4.1 Weiszfeld Algorithm

The algorithm 2 requires a line search to determine the step length in the descending gradient direction. It is possible to give a closed-form step length that still guarantees convergence. The algorithm is derived from the classical Weiszfeld algorithm [81] that finds the geometric median (L_1-mean) of points in $\mathbb{R}^n$. The application of the Weiszfeld algorithm to the geodesic L_1 averaging problem was shown in [27]. It differs from algorithm 2 only in the method of determining the step length, s.

The Weiszfeld algorithm in $\mathbb{R}^n$ is a gradient descent algorithm. Given points $\mathbf{x}_i \in \mathbb{R}^n$, the downhill gradient of the cost at a given point $\mathbf{y}$ is the sum of the unit vectors directed from $\mathbf{y}$ to each of the points $\mathbf{x}_i$. Given a current estimate $\mathbf{y}^t$, the next estimate of the minimum in a gradient descent algorithm is given by

$$y^{t+1} = y^t + \lambda \sum_{i=1}^n \frac{\mathbf{x}_i - \mathbf{y}^t}{\|\mathbf{x}_i - \mathbf{y}^t\|}\ ,$$

where $\lambda > 0$ is some value controlling the step-size along the gradient direction. The choice of step size in the Weiszfeld algorithm is set to be $\lambda = \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{y}^t\|^{-1}$. Writing $s_i^t = \|\mathbf{x}_i - \mathbf{y}^t\|$, we then find that

$$\mathbf{y}^{t+1} = \mathbf{y}^t + \frac{\sum_{i=1}^n (\mathbf{x}_i - \mathbf{y}^t)/s_i^t}{\sum_{i=1}^n 1/s_i^t} \tag{19}$$

$$= \frac{\sum_{i=1}^n \mathbf{x}_i/s_i^t}{\sum_{i=1}^n 1/s_i^t}\ . \tag{20}$$

As long as the intermediate iterates $\mathbf{y}_i^t$ do not concide with any of the points $\mathbf{x}_i$, this algorithm will provably converge to the geometric median of the points [81]. Convergence may not be fast. If in fact $\mathbf{y}^t$ coincides with some point $\mathbf{x}_i$, then the algorithm will get stuck at this point. A simple strategy in this case is to displace the iterate $\mathbf{y}^t$ slightly and continue. It may easily be shown that successive iterates will "escape" from some point $\mathbf{x}_i$, not the minimum, by approximately doubling the distance at each iteration.

Geodesic median in $\mathrm{SO}(3)$**.** We now consider the problem of computing the L_1 geodesic mean in $\mathrm{SO}(3)$. To be able to apply the Weiszfeld algorithm, we transition back and forth between $\mathrm{SO}(3)$, and its tangent space centred at the current estimate. The following lemma shows that we can use the geometric median in the tangent space to find the geodesic median in rotation space $\mathrm{SO}(3)$.

Lemma 3 *If $\mathtt{S}$ is the geodesic median of rotations $\mathtt{R}_i$, then the origin $\mathbf{0} \in \mathbb{R}^3$ is the geometric median of the points $\log_{\mathtt{S}}(\mathtt{R}_i)$. Conversely, if $\mathtt{S}$ and all $\mathtt{R}_i$ lie in a convex set in $\mathrm{SO}(3)$, and $\mathbf{0}$ is the geometric median of the points $\log_{\mathtt{S}}(\mathtt{R}_i)$, then $\mathtt{S}$ is the geodesic median of the $\mathtt{R}_i$.*

Proof. If $\mathtt{S}$ is the geodesic median of the $\mathtt{R}_i$, then the gradient of the cost function, $C(\mathtt{S}) = \sum_{i=1}^n d_\angle(\mathtt{R}_i, \mathtt{S})$ is zero at $\mathtt{S}$. The gradient is defined as a vector in the tangent space at $\mathtt{S}$, which may be computed to be the vector

$$\nabla_C = -\sum_{i=1}^n \frac{\log_{\mathtt{S}}(\mathtt{R}_i)}{\|\log_{\mathtt{S}}(\mathtt{R}_i)\|} \ .$$

Writing $\mathbf{r}_i = \log_{\mathtt{S}}(\mathtt{R}_i)$, we see that this gradient vector is also the gradient of the cost $C'(\mathbf{s}) = \sum_{i=1}^n \|\mathbf{r}_i - \mathbf{s}\|$ at $\mathbf{s} = \mathbf{0} \in \mathbb{R}^3$. In other words, $\mathbf{s} = \mathbf{0}$ is a critical point of C'. However, since this cost function is convex in $\mathbb{R}^3$, it follows that $\mathbf{s} = \mathbf{0}$ is a minimum of the cost function, and so $\mathbf{s} = \mathbf{0}$ is the geometric median of the $\mathbf{r}_i$ in $\mathbb{R}^3$.

On the other hand, if $\mathbf{s} = \mathbf{0}$ is the minimum of $C'(\mathbf{s})$, then by the same argument the gradient of C is zero at $\mathtt{S}$, so $\mathtt{S}$ is at least a critical point of the cost. However, it is not true that the cost function C is convex, or has a single minimum on the whole of $\mathrm{SO}(3)$. For this reason, we need the condition that all $\mathtt{R}_i$ lie in a convex set (theorem 5). In this case, the cost C is convex in the set and the global minimum of C lies within the ball, and hence at $\mathtt{S}$. □

Given this lemma, we are led to propose an algorithm for finding the geodesic median in $\mathrm{SO}(3)$, based on the Weiszfeld algorithm on the tangent space. Given rotations $\mathtt{R}_i \in \mathrm{SO}(3)$, we proceed as follows.

1. Find an initial estimate $\mathtt{S}^0$ for the median. Such an estimate may already be know, or else we may take the L_2-mean of the rotations $\mathtt{R}_i$ as a starting point.
2. At any time $t = 0, 1, \ldots$ apply the logarithm map centred at $\mathtt{S}^t$, to compute $\mathbf{v}_i = \log_{\mathtt{S}^t}(\mathtt{R}_i)$.
3. (Weiszfeld step): Set
$$\delta = \frac{\sum_{i=1}^n \mathbf{v}_i/\|\mathbf{v}_i\|}{\sum_{i=1}^n 1/\|\mathbf{v}_i\|}$$
4. Set $\mathtt{S}^{t+1} = \exp(\delta)\mathtt{S}^t$.
5. Repeat steps 2 to 4 until convergence.

Thus, this algorithm may be thought of as carrying out successive iterative steps of the Weiszfeld algorithm, each step taking place in the tangent space centred at the current estimate.

For computational efficiency, it is simpler to work with the quaternion representations $\mathbf{r}_i$ of the rotations $\mathtt{R}_i$, since mapping between quaternions and angle-axis representation is simpler than computing the exponential and logarithm maps. In addition quaternion multiplication is faster than matrix multiplication. Let Q be the unit quaternions. The mapping $q : \mathbb{R}^3 \to Q$ given by

$$q : \theta\hat{\mathbf{v}} \mapsto (\cos(\theta/2), \sin(\theta/2)\hat{\mathbf{v}})$$

maps between angle-axis and quaternion representation of a rotation. Then steps 2 to 4 of the above algorithm are replaced by

$$\theta_i\hat{\mathbf{v}}_i = q^{-1}(\mathbf{r}_i \cdot \bar{\mathbf{s}}^t)$$

$$\delta = \frac{\sum_{i=1}^n \hat{\mathbf{v}}_i}{\sum_{i=1}^n 1/\theta_i}$$

$$\mathbf{s}^{t+1} = q(\delta) \cdot \mathbf{s}^t$$

where $\bar{\mathbf{s}}^t$ represents the conjugate (inverse) of the quaternion $\mathbf{s}^t$. A further alternative is to use the Baker-Campbell-Hausdorff formula (see e.g. [23]) to work entirely in angle-axis space, but this is essentially equivalent to this use of quaternions.

As is shown in lemma 3, a stationary point of this algorithm (for which $\mathtt{S}^t = \mathtt{S}^{t+1}$) must be the geodesic median of the rotations $\mathtt{R}_i$, provided that $d_\angle(\mathtt{S}^t, \mathtt{R}_i) < \pi/2$ for all i.

The Euclidean metric in a tangent space is related within constant bounds to the angular metric in $\mathrm{SO}(3)$, so it is plausible that this algorithm will converge. However, convergence of this algorithm follows from the results of [20, 2]. More precisely, it was shown in [20] that if all the $\mathtt{R}_i$ lie within a ball of radius $\pi/4$, the above algorithm converges to the so-called solipsistic mean (the minimum of the cost function within the given ball) provided that (1) not all the $\mathtt{R}_i$ lie on a single geodesic, and (2) the algorithm does not step outside that ball. Condition (2) can be shown to hold in this setting. Alternatively, restriction (2) can be overcome using step size control and projection techniques [83], although this negates the conceptual simplicity of the Weiszfeld algorithm. Finally, it was shown in [2] that the solipsistic mean is in fact the global mean if all the $\mathtt{R}_i$ lie within a ball of radius $\pi/4$.

5.5 The Chordal L_2-mean

The cost function for rotation averaging under the L_2 chordal metric is

$$C(\mathtt{R}) = \sum_{i=1}^n d_{\mathrm{chord}}(\mathtt{R}_i, \mathtt{R})^2 \ , \tag{21}$$

and the chordal L_2-mean of a set of rotations $\mathtt{R}_i$ is defined as the rotation that minimizes this cost. It is usually called the *projected* or *induced arithmetic mean* [57,72]. As shown in fig 2, the chordal distance metric is not convex beyond a ball of radius $\pi/2$. Thus the squared chordal distance has substantially different convexity properties compared to the squared geodesic distance (theorem 11). Making the appropriate changes we get the following analogue to corollary 1.

Theorem 7 *Let $\mathtt{R}_i$ be rotations satisfying $d_\angle(\mathtt{R}_i, \mathtt{S}) < \pi/4$ for some rotation $\mathtt{S}$ and for all i, then*

$$C(\mathtt{R}) = \sum_{i=1}^{n} d_{\text{chord}}(\mathtt{R}_i, \mathtt{R})^2$$

is strictly convex on $B(\mathtt{S}, \pi/4)$, and hence has a single isolated minimum on that set.

Theorem 5, specifying possible locations of global minima, applies unchanged and hence we have the following global result.

Corollary 2 *Let $\mathtt{R}_i$ be rotations lying in a convex set B of radius less than $\pi/4$, then the unique global chordal L_2-mean lies in B and moreover the cost function $C(\mathtt{R}) = \sum_{i=1}^{n} d_{\text{chord}}(\mathtt{R}_i, \mathtt{R})^2$ is strictly convex on some ball $B(\mathtt{S}, \pi/4) \supset B$.*

There is no direct analogue of theorem 3 for the chordal L_2-mean.

A closed-form algorithm. Given this seemingly less favourable convexity situation, compared with the geodesic and quaternion means, it is perhaps surprising that there is a closed-form algorithm for finding the global minimum of (21). The solution given in [54] involves the quaternion representation of the rotations. Let rotations $\mathtt{R}_i$ be given, and let $\mathbf{r}_i$ be chosen quaternion representations. Form the matrix $\mathtt{A} = \sum_{i=1}^{n} \mathbf{r}_i \mathbf{r}_i{}^\top$, which is a 4×4 symmetric matrix. Note that it does not depend on the choice between $\mathbf{r}$ and $-\mathbf{r}$. Now, let $\mathbf{s}^*$ be the eigenvector of $\mathtt{A}$ corresponding to the maximum eigenvalue. We claim that $\mathbf{s}^*$ is the quaternion representation for the minimum of the cost (21).

Let $\mathbf{s}$ be a quaternion, and $\cos(\alpha_i) = \langle \mathbf{r}_i, \mathbf{s} \rangle$. Then

$$\mathbf{s}^\top \mathtt{A} \mathbf{s} = \sum_{i=1}^{n} \cos^2(\alpha_i) = \sum_{i=1}^{n} \cos^2(\theta_i/2),$$

where $\theta_i = d_\angle(\mathtt{R}_i, \mathtt{S})$, cf. section 4. Then $\mathbf{s}^*$ is the vector that maximizes the left-hand side of this equation. Thus,

$$\begin{aligned} \mathbf{s}^* &= \operatorname*{argmax}_{\mathbf{s}} \sum_{i=1}^{n} \cos^2(\theta_i/2) \\ &= \operatorname*{argmin}_{\mathbf{s}} \sum_{i=1}^{n} \sin^2(\theta_i/2) \\ &= \operatorname*{argmin}_{\mathbf{s}} \sum_{i=1}^{n} d_{\text{chord}}(\mathtt{R}_i, \mathtt{S})^2 \ . \end{aligned}$$

Note that by using quaternions, we obtain the chordal mean, not the quaternion mean. This algorithm will fail to give a unique solution only when the matrix $\mathtt{A}$ has repeated maximum eigenvalues.

Closed form using rotations. A full characterization of all the minima of the cost function (21) can also be given in terms of the matrix representations of the rotations [72]. Let

$$\mathtt{C}_\text{e} = \sum_{i=1}^{n} \mathtt{R}_i \in \mathbb{R}^{3\times 3}$$

Let $\langle \cdot, \cdot \rangle$ represent the Frobenius inner product (sum of the elementwise products of two matrices). Then, if $\mathtt{R}_i$ and $\mathtt{S}$ are rotations,

$$\begin{aligned} \sum_{i=1}^{n} d_{\text{chord}}(\mathtt{R}_i, \mathtt{S})^2 &= \sum_{i=1}^{n} \|\mathtt{R}_i - \mathtt{S}\|_\text{F}^2 \\ &= \sum_{i=1}^{n} \langle \mathtt{R}_i - \mathtt{S}, \mathtt{R}_i - \mathtt{S} \rangle \\ &= \sum_{i=1}^{n} \left(\langle \mathtt{R}_i, \mathtt{R}_i \rangle - 2 \langle \mathtt{R}_i, \mathtt{S} \rangle + \langle \mathtt{S}, \mathtt{S} \rangle \right) \\ &= K - 2 \langle \mathtt{C}_\text{e}, \mathtt{S} \rangle \ , \end{aligned}$$

where K is a constant (independent of $\mathtt{S}$). Therefore,

$$\begin{aligned} \operatorname*{argmin}_{\mathtt{S}\in SO(3)} \sum_{i=1}^{n} d_{\text{chord}}(\mathtt{R}_i, \mathtt{S})^2 &= \operatorname*{argmax}_{\mathtt{S}\in SO(3)} \langle \mathtt{C}_\text{e}, \mathtt{S} \rangle \\ &= \operatorname*{argmin}_{\mathtt{S}\in SO(3)} \|\mathtt{C}_\text{e} - \mathtt{S}\|_\text{F} \end{aligned}$$

Thus minimizing the L_2 chordal cost function is equivalent to finding the closest matrix $\mathtt{S}$ to $\mathtt{C}_\text{e}$ under the Frobenius norm.

The matrix $\mathtt{S}$ that we seek is obtained using the Singular Value Decomposition. Let $\mathtt{C}_\text{e} = \mathtt{U}\,\mathtt{D}\,\mathtt{V}^\top$ where the diagonal elements of $\mathtt{D}$ are arranged in descending order. If $\det(\mathtt{U}\mathtt{V}^\top) \geq 0$, then set $\mathtt{S} = \mathtt{U}\mathtt{V}^\top$. Otherwise set $\mathtt{S} = \mathtt{U}\,\text{diag}(1, 1, -1)\mathtt{V}^\top$. The matrix $\mathtt{S}$ so obtained is the closest rotation to $\mathtt{C}_\text{e}$, and hence the required rotation minimizing (21).

5.6 The Chordal L_1-mean

The chordal L_1 mean of a set of rotations $\mathtt{R}_i$ is defined as the minimum of

$$C(\mathtt{R}) = \sum_{i=1}^{n} d_{\text{chord}}(\mathtt{R}_i, \mathtt{R}) = \sum_{i=1}^{n} 2\sqrt{2}\, \sin(\theta_i/2)$$

where $\theta_i = d_\angle(\mathtt{R}_i, \mathtt{R})$ denotes the angle of the rotation $\mathtt{R}_i\mathtt{R}^\top$.

Although the chordal distance is not convex (theorem 11), theorem 5 still applies, constraining the possible global minima in case the $\mathtt{R}_i$ lie in a convex set. However, because of non-convexity, we can not assert that multiple global minima do not exist in this case. In fact, when $n = 2$, or when the rotations all lie on or even near a single geodesic, it is easy to find cases where multiple local minima exist, centred on the individual rotations.

Since the L_1 metric is not differentiable for $\mathtt{R} = \mathtt{R}_i$, the shape of the cost function $C(\mathtt{R})$ is a little complex. Nevertheless, we can easily compute the gradient

$$\nabla C(\mathtt{R}) = -\sqrt{2} \cdot \mathtt{R} \sum_{i=1}^{n} \log\left(\mathtt{R}^\top \mathtt{R}_i\right) \frac{\cos(\theta_i/2)}{\theta_i} \ ,$$

see the appendix for the details. This formula can be viewed as a weighted version of the gradient for the geodesic L_1-mean where the weights are $\sqrt{2}\,\cos(\theta_i/2)$.

We propose a Riemannian gradient descent algorithm with geodesic line search to compute the chordal L_1-mean, or at least a critical point of the cost function $C(\mathtt{R})$.

Choose a tolerance $\varepsilon > 0$.
Set $\mathtt{R} := d^2_{\text{chord}}\text{-mean}(\{\mathtt{R}_1, \ldots, \mathtt{R}_n\})$.
loop
  Compute $\mathbf{r} := \sqrt{2}\sum_{i=1}^n \log(\mathtt{R}^\top \mathtt{R}_i)\cos(\theta_i/2)/\theta_i$.
  Compute $s^* := \text{argmin}_{s \geq 0}\ C(\mathtt{R}\exp(s\mathbf{r}))$.
  if $\|s^*\mathbf{r}\| < \varepsilon$ **then**
    return $\mathtt{R}$
  end if
  Update $\mathtt{R} := \mathtt{R}\exp(s^*\mathbf{r})$.
end loop

Algorithm 3: Computing the chordal L_1-mean on $\mathrm{SO}(3)$

As long as we avoid the points of non-differentiability, this algorithm should converge, at least to a local minimum.

5.7 The Quaternion L_2-mean

The quaternion L_2-mean of a set of rotations $\mathtt{R}_i$ is defined as $\text{argmin}_{\mathtt{R}\in\mathrm{SO}(3)} \sum_{i=1}^n d_{\text{quat}}(\mathtt{R}_i, \mathtt{R})^2$. Since the squared quaternion distance enjoys the same convexity properties as the squared angular distance (theorem 11), applying the previous theorems we get the following strong global existence and uniqueness result.

Theorem 8 *Let $\mathtt{R}_i$ be rotations lying in a convex set B of radius less than $\pi/2$, then the unique global quaternion L_2-mean lies in B and moreover the cost function $C(\mathtt{R}) = \sum_{i=1}^n d_{\text{quat}}(\mathtt{R}_i, \mathtt{R})^2$ is strictly convex on some ball $B(\mathtt{S}, \pi/2) \supset B$. In the general case, $C(\mathtt{R})$ is strictly convex, except on the union of sets*

$$\Pi_i = \{\mathtt{S} \in \mathrm{SO}(3) \,|\, d_\angle(\mathtt{R}_i, \mathtt{S}) = \pi\}.$$

It is non-differentiable on the union of the sets Π_i, and has at most one minimum on each of the $\binom{n}{3} + n$ closed regions bounded by the Π_i.

The following theorem shows how the quaternion L_2-mean may be computed.

Theorem 9 *Let $\mathtt{R}_i$ be rotations satisfying $d_\angle(\mathtt{R}_i, \mathtt{S}) < \pi/2$ for some rotation $\mathtt{S}$ and for all i. Let $\mathbf{s}$ be a quaternion representation of $\mathtt{S}$ and let $\mathbf{r}_i$ be the quaternion representation of $\mathtt{R}_i$ chosen with sign such that $\|\mathbf{r}_i - \mathbf{s}\|_2$ is the smaller of the two choices. Then the quaternion L_2-mean of the rotations $\mathtt{R}_i$ is represented by the quaternion $\bar{\mathbf{r}}/\|\bar{\mathbf{r}}\|$, where $\bar{\mathbf{r}} = \sum_{i=1}^n \mathbf{r}_i$.*

Proof. Let $\mathtt{T}$ be a rotation and $\mathbf{t}$ be a quaternion representation. The quaternion distance to a rotation $\mathtt{R}_i$ is given by $\|\mathbf{t} - \mathbf{r}_i\|_2$ or $\|\mathbf{t} + \mathbf{r}_i\|_2$, whichever is smaller. Thus, the mean of the rotations $\mathtt{R}_i$ is given by the quaternion $\mathbf{t}$ that minimizes

$$\sum_{i=1}^n \|\mathbf{t} - \varepsilon_i \mathbf{r}_i\|_2^2$$

over $\mathbf{t}$ and all choices of $\varepsilon_i = \pm 1$. First, let us assume that this minimum is achieved when all $\varepsilon_i = 1$. Let α_i equal the angle between $\mathbf{t}$ and $\mathbf{r}_i$ as vectors in $\mathbb{R}^4$. Then, the quaternion mean is found by minimizing

$$\sum_{i=1}^n \|\mathbf{t} - \mathbf{r}_i\|_2^2 = \sum_{i=1}^n 4\sin^2(\alpha_i/2) = \sum_{i=1}^n 2(1 - \cos(\alpha_i)) \ .$$

This is equivalent to maximizing

$$\sum_{i=1}^n \cos(\alpha_i) = \sum_{i=1}^n \langle \mathbf{t}, \mathbf{r}_i \rangle = \left\langle \mathbf{t}, \sum_{i=1}^n \mathbf{r}_i \right\rangle \ ,$$

where $\langle \mathbf{t}, \mathbf{r}_i \rangle$ represents the inner product of $\mathbf{t}$ and $\mathbf{r}_i$ as vectors in $\mathbb{R}^4$. However, since $\mathbf{t}$ must be a unit vector, this quantity is clearly maximized by setting $\hat{\mathbf{t}} = \sum_{i=1}^n \mathbf{r}_i$ and $\mathbf{t} = \hat{\mathbf{t}}/\|\hat{\mathbf{t}}\|_2$. Thus, we have proved the required result, under the assumption that all the signs ε_i were positive. Denote the vector $\hat{\mathbf{t}}$ defined in this way as $\hat{\mathbf{t}}_0$ and note that $\|\hat{\mathbf{t}}_0\|_2^2$ equals the associated sum of angle cosines $\cos(\alpha_i)$.

Now, assume that some ε_i are negative, and so

$$\sum_{i=1}^n d_{\text{quat}}(\mathtt{T}, \mathtt{R}_i)^2 = \sum_{i \in S^+} \|\mathbf{t} - \mathbf{r}_i\|^2 + \sum_{i \in S^-} \|\mathbf{t} + \mathbf{r}_i\|^2$$

where S^+ and S^- are the corresponding division of $\{1, \ldots, n\}$ into two parts. By the same argument as before, this quantity is maximized with respect to $\mathbf{t}$ by setting

$$\hat{\mathbf{t}}_1 = \sum_{i \in S+} \mathbf{r}_i - \sum_{i \in S^-} \mathbf{r}_i = \mathbf{r}^+ - \mathbf{r}^- \quad \text{and} \quad \mathbf{t} = \hat{\mathbf{t}}_1/\|\hat{\mathbf{t}}_1\|_2 \ ,$$

where $\mathbf{r}^+$ and $-\mathbf{r}^-$ are the resultants of the two sets of vectors. On the other hand, the original $\hat{\mathbf{t}}_0 = \mathbf{r}^+ + \mathbf{r}^-$. The proof will be completed by showing that $\|\mathbf{r}^+ + \mathbf{r}^-\|_2 > \|\mathbf{r}^+ - \mathbf{r}^-\|_2$, for then $\|\hat{\mathbf{t}}_0\|_2^2 > \|\hat{\mathbf{t}}_1\|_2^2$ and the sum of angle cosines is maximized when all the $\varepsilon_i = 1$.

Now, since each $\mathbf{r}_i$ lies within an angle $\pi/4$ of $\mathbf{s}$ (remember $\mathbf{s}$) by hypothesis, so must both $\mathbf{r}^+$ and $\mathbf{r}^-$. This means that $\mathbf{r}^+$ and $\mathbf{r}^-$ lie within an angle of $\pi/2$ of each other. On the other hand, $\mathbf{r}^+$ and $-\mathbf{r}^-$ differ in direction by more than $\pi/2$. It follows that $\|\mathbf{r}^+ + \mathbf{r}^-\|_2 > \|\mathbf{r}^+ - \mathbf{r}^-\|_2$, and the proof is complete. □

5.8 The Quaternion L_1-mean

The quaternion L_1-mean is defined as the minimum of the cost function

$$C(\mathtt{R}) = \sum_{i=1}^{n} d_{\text{quat}}(\mathtt{R}_i, \mathtt{R}) = 2\sum_{i=1}^{n} \sin(\theta_i/4)$$

where $\theta_i = d_\angle(\mathtt{R}_i, \mathtt{R})$ denotes the angle of the rotation $\mathtt{R}_i\mathtt{R}^\top$. We have

$$\nabla\, C(\mathtt{R}) = -\frac{1}{2}\mathtt{R}\sum_{i=1}^{n} \log\left(\mathtt{R}^\top\mathtt{R}_i\right)\frac{\cos(\theta_i/4)}{\theta_i}.$$

As with the chordal L_1-mean, we do not have any uniqueness result due to a lack of convexity. The best we can offer is again a Riemannian gradient descent algorithm with geodesic line search to compute critical points of $C(\mathtt{R})$. We leave it to the reader to make the obvious modifications to algorithm 3.

6 The Conjugate Rotation Averaging Problem

The general form of the conjugate averaging problem is as follows. Let $(\mathtt{R}_i, \mathtt{L}_i); i = 1, \ldots, n$ be pairs of rotations. The conjugate averaging problem is to find the rotation $\mathtt{S}$ that minimizes

$$C(\mathtt{S}) = \sum_{i=1}^{n} d^p(\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}, \mathtt{L}_i)\ . \qquad (22)$$

The motivation for this problem is that we may have estimates $\mathtt{L}_i$ and $\mathtt{R}_i$ of the motion of left and right cameras expressed in different coordinate frames, local to the two cameras. We wish to find the rotation of one coordinate frame to the other. Under noise-free conditions, the relationship is $\mathtt{L}_i = \mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}$, where $\mathtt{S}$ expresses the rotation of the right coordinate frame with respect to the left one.

Under different distance metrics, this problem has different solutions. In this section we will give algorithms for some of the various metrics discussed before.

Minimal configurations for conjugate averaging. The first question is, how many pairs $(\mathtt{R}_i, \mathtt{L}_i)$ are needed in order to estimate $\mathtt{S}$.

If only one rotation pair $(\mathtt{R}, \mathtt{L})$ is given, then there is not a unique solution. Let $\mathtt{S}^*$ be a rotation that minimizes $d^p(\mathtt{S}^{-1}\mathtt{R}\mathtt{S}, \mathtt{L})$ and define $\mathtt{S}(t) = \exp[t\mathbf{r}]_\times$, where $\mathbf{r}$ is the axis of rotation of $\mathtt{R}$. Then $\mathtt{S}(t)$ commutes with $\mathtt{R}$, so $d^p(\mathtt{S}^{*-1}\mathtt{S}(t)^{-1}\mathtt{R}\mathtt{S}(t)\mathtt{S}^*, \mathtt{L}) = d^p(\mathtt{S}^{*-1}\mathtt{R}\mathtt{S}^*, \mathtt{L})$ for all t. Consequently $\mathtt{S}(t)\mathtt{S}^*$ is also a minimizer of the cost function for all t. The rotations that minimize the cost lie along the geodesic $\exp[t\mathbf{r}]_\times\mathtt{S}^*$.

If there are two rotation pairs, then the optimum must lie on the intersection of two geodesics in general, and these intersect at a single point. Hence, in general, two rotation pairs are sufficient to give a unique solution, unless the rotations $\mathtt{R}_i$ are about the same axis.

Alignment of rotation axes. As we shall see, the solution to the conjugate rotation averaging problem is closely related to alignment of the axes of the rotations. Thus, let $\hat{\mathbf{r}}_i$ and $\hat{\mathbf{l}}_i$ be the rotation axes of the rotations, then we may consider the problem of finding a rotation $\mathtt{S}$ such that $\mathtt{S}\hat{\mathbf{l}}_i = \hat{\mathbf{r}}_i$. An optimal rotation to solve this problem is given in [33].

There is, however, a difficulty with this approach, namely the ambiguity between a rotation axis and the oppositely directed axis, between $\hat{\mathbf{r}}_i$ and $-\hat{\mathbf{r}}_i$. A rotation may be represented by a rotation through an angle θ about an axis $\hat{\mathbf{r}}$ or as a rotation through an angle $2\pi - \theta$ about the opposite axis $-\hat{\mathbf{r}}$. One may resolve this issue by choosing the rotation angle to not exceed π. However, there are still two choices of the axis in the case when the rotation is through an angle π. In addition, in the case of error in the measurement of rotations through an angle close to π, the wrong axis may be chosen. In this case no rotation will closely align the rotation axes $\hat{\mathbf{r}}_i$ and $\hat{\mathbf{l}}_i$.

This ambiguity may be resolved under certain reasonable conditions.

1. There exists a value $\theta_{\max} < \pi$ such that $\angle(\mathtt{R}_i) \leq \theta_{\max}$ and $\angle(\mathtt{L}_i) \leq \theta_{\max}$ for all i.
2. For the "true" solution $\mathtt{S}$ being sought, the maximum error for any of the rotations $\mathtt{R}_i$ is $\delta_{\max}$. Thus, $d_\angle(\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}, L_i) \leq \delta_{\max}$ when $\mathtt{S}$ is the required solution. This condition is reasonable if we assume that the rotations $\mathtt{R}_i$ and $\mathtt{L}_i$ are all measured with a maximum angle error of $\delta_{\max}/2$.
3. $\theta_{\max} + \delta_{\max}/2 \leq \pi$.

Thus, we are assuming that the errors plus angles are not too large. In particular, since $\delta_{\max} \leq \pi$ we see that the last two conditions always hold if $\theta_{\max} \leq \pi/2$.

For the application we are interested in, where $\mathtt{R}_i$ and $\mathtt{L}_i$ are relative rotations between two positions of a camera, the rotation angle of $\mathtt{R}_i$ can not be very large. If for instance the rotation $\mathtt{R}$ between two positions of a camera approaches π, then at least for normal cameras, there will be no points visible in both images, and hence no way to estimate the rotation $\mathtt{R}$. Normally, the rotation $\mathtt{R}$ between two positions of the camera will not exceed the field of view of the camera, otherwise there will not be any matched points for the two camera views (except possibly for points lying between the two camera positions).

We make the observation that if $\mathtt{R}_i$ and $\mathtt{L}_i$ are exactly conjugate, that is, $\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S} = \mathtt{L}_i$ for some rotation $\mathtt{S}$, then they have the same rotation angle. Under the conditions just given, the angles $\angle(\mathtt{R}_i)$ and $\angle(\mathtt{L}_i)$ can not differ by more than $2\delta_{\max}$.

Lemma 4 *Let* $\mathbf{r} = (\cos(\theta_1/2), \hat{\mathbf{r}}\sin(\theta_1/2))$ *and* $\mathbf{l} = (\cos(\theta_2/2), \hat{\mathbf{l}}\sin(\theta_2/2))$ *be quaternions representing rotations* $\mathtt{R}$ *and* $\mathtt{L}$*, with* $\theta_i \leq \theta_{\max} < \pi$*, for* $i = 1, 2$ *(meaning that* $\mathbf{r}$ *and* $\mathbf{l}$ *lie in the upper unit half sphere). If* $\mathtt{S}$ *is a rotation satisfying the constraint*

$$d_\angle(\mathtt{S}^{-1}\mathtt{RS}, \mathtt{L}) \leq 2\sin((\pi - \theta_{\max})/2)$$

and $\mathbf{s}$ *is either of its two quaternion representations, then* $\|\mathbf{r}\cdot\mathbf{s} - \mathbf{s}\cdot\mathbf{l}\|_2 \leq \|\mathbf{r}\cdot\mathbf{s} + \mathbf{s}\cdot\mathbf{l}\|_2$*, and so* $d_{\text{quat}}(\mathtt{S}^{-1}\mathtt{RS}, L) = \|\mathbf{r}\cdot\mathbf{s} - \mathbf{s}\cdot\mathbf{l}\|_2$.

Proof. Observe that if $\mathbf{r} = (r_0, \mathbf{r}')$, where $\mathbf{r}'$ is a 3-vector in the direction of the rotation axis, then $\mathbf{s}^{-1}\cdot\mathbf{r}\cdot\mathbf{s} = (r_0, \mathtt{S}^{-1}\mathbf{r}')$. Thus, conjugating by $\mathbf{s}^{-1}$ does not change the first component of the quaternion, and rotates the axis by $\mathtt{S}^{-1}$. However, if θ is the rotation angle of $\mathtt{R}$, then by hypothesis, $r_0 = \cos(\theta/2) \geq \cos(\theta_{\max}/2)$. Similarly for $\mathbf{l}$, we have $l_0 \geq \cos(\theta_{\max}/2)$. Therefore,

$$\begin{aligned}\|\mathbf{s}^{-1}\cdot\mathbf{r}\cdot\mathbf{s} + \mathbf{l}\|_2 &\geq 2\cos(\theta_{\max}/2)\\ &= 2\sin((\pi - \theta_{\max})/2)\ .\end{aligned}$$

On the other hand, by hypothesis, $d_{\text{quat}}(\mathtt{S}^{-1}\mathtt{RS}, \mathtt{L}) \leq 2\sin((\pi - \theta_{\max})/2)$. It follows that $\|\mathbf{r}\cdot\mathbf{s} - \mathbf{s}\cdot\mathbf{l}\|_2 \leq \|\mathbf{r}\cdot\mathbf{s} + \mathbf{s}\cdot\mathbf{l}\|_2$, and $d_{\text{quat}}(\mathtt{S}^{-1}\mathtt{RS}, \mathtt{L}) = \|\mathbf{r}\cdot\mathbf{s} - \mathbf{s}\cdot\mathbf{l}\|_2$, as we wished to prove. □

6.1 The Quaternion L_2-mean for Conjugate Averaging

The squared quaternion distance seems to be best suited to this particular averaging problem. We give here an optimal solution for the squared quaternion distance under the conditions 1 – 3.

Under these conditions, we can modify the optimization problem slightly to restrict the solution $\mathtt{S}$ so that the errors are bounded in this way. Thus, our modified problem is

$$\begin{aligned}&\textbf{Minimize} \quad C(\mathtt{S}) = \sum_{i=1}^{n} d_{\text{quat}}(\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}, \mathtt{L}_i)^2\\ &\textbf{Subject to} \quad d_\angle(\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}, \mathtt{L}_i) \leq \delta_{\max} \text{ for all } i\ .\end{aligned} \tag{23}$$

where $\delta_{\max} \leq 2(\pi - \theta_{\max})$.

Note that this condition may be written in terms of the quaternion metric as

$$d_{\text{quat}}(\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}, \mathtt{L}_i) \leq 2\sin((\pi - \theta_{\max})/2)\ .$$

The purpose of this condition is to allow us to remove the sign ambiguity about the quaternion representation of a rotation and the quaternion metric.

A linear solution. We now outline a linear algorithm for estimating the matrix $\mathtt{S}$, under the squared quaternion distance. Let $\mathbf{r}_i$ and $\mathbf{l}_i$ be quaternion representations of the rotations $\mathtt{R}_i$ and $\mathtt{L}_i$, chosen such that $\mathbf{r}_i = (\cos(\theta_i/2), \hat{\mathbf{r}}\sin(\theta_i/2))$ with $\theta_i < \pi$. This means that the first component $\cos(\theta_i/2)$ of the quaternion is positive. This fixes the choice between $\mathbf{r}_i$ and $-\mathbf{r}_i$. We define $\mathbf{l}_i$ similarly.

Now, consider the equation $\mathtt{R}_i\mathtt{S} = \mathtt{SL}_i$, and write it in terms of quaternions as $\mathbf{r}_i\cdot\mathbf{s} - \mathbf{s}\cdot\mathbf{l}_i = \mathbf{0}$. As before, $\cdot$ represents quaternion multiplication. Since quaternion multiplication is bilinear in terms of the entries of the two quaternions involved, this gives a homogeneous linear equation in terms of the entries of $\mathbf{s}$. Stacking all these equations into one and finding the least squares solution such that $\|\mathbf{s}\|_2 = 1$, we may solve for $\mathbf{s}$. This gives a simple linear way to solve this problem. Under the conditions stated above, we can prove that this algorithm finds the global minimum with respect to the squared quaternion distance as follows.

The question is, what does this linear solution represent when the equations $\mathbf{r}_i\cdot\mathbf{s} - \mathbf{s}\cdot\mathbf{l}_i = \mathbf{0}$ are not exactly satisfied. The least-squares solution to a set of such equations will find $\mathbf{s}$ that minimizes $\sum_{i=1}^n \|\delta_i\|_2^2$, where $\delta_i = \mathbf{r}_i\cdot\mathbf{s} - \mathbf{s}\cdot\mathbf{l}_i$. Thus, the linear solution will minimize

$$\sum_{i=1}^{n} \|\mathbf{r}_i\cdot\mathbf{s} - \mathbf{s}\cdot\mathbf{l}_i\|_2^2 = \sum_{i=1}^{n} d_{\text{quat}}^2(\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}, \mathtt{L}_i)\ .$$

We have used lemma 4 in this last step.

Aligning the axes. Solving this problem under the L_2 quaternion metric is equivalent to simply aligning the rotation axes, appropriately weighted. This gives a slightly different algorithm, as follows.

Let the rotations $\mathtt{R}_i$ and $\mathtt{L}_i$ be represented by the quaternions

$$\mathbf{r}_i = (\cos(\theta_i/2), \hat{\mathbf{r}}_i\sin(\theta_i/2))$$

and

$$\mathbf{l}_i = (\cos(\phi_i/2), \hat{\mathbf{l}}_i\sin(\phi_i/2)),$$

respectively. These quaternions are chosen such that the rotation angles θ_i and ϕ_i are less than π. As observed previously, the quaternion corresponding to $\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}$ is

$$\mathbf{s}^{-1}\cdot\mathbf{r}_i\cdot\mathbf{s} = (\cos(\theta_i/2), \mathtt{S}^{-1}\hat{\mathbf{r}}_i\sin(\theta_i/2))\ .$$

As we showed above, minimizing $\sum_i d_{\text{quat}}(\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}, \mathtt{L}_i)^2$ under the constraint that $d_{\text{quat}}(\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}, \mathtt{L}_i)^2 \leq 2(\pi - \theta_{\max})$ is equivalent to minimizing $\sum_i \|\mathbf{r}_i\cdot\mathbf{s} - \mathbf{s}\cdot\mathbf{l}_i\|_2^2$. Now, it is easily observed that

$$\begin{aligned}\sum_i \|\mathbf{r}_i\cdot\mathbf{s} - \mathbf{s}\cdot\mathbf{l}_i\|_2^2 &= \sum_i \|\mathbf{s}^{-1}\cdot\mathbf{r}_i\cdot\mathbf{s} - \mathbf{l}_i\|_2^2 =\\ &\sum_i \|\mathtt{S}^{-1}\hat{\mathbf{r}}_i\sin(\theta_i/2) - \hat{\mathbf{l}}_i\sin(\phi_i/2)\|_2^2 + K\end{aligned}$$

where $K = \sum_i \left(\cos(\theta_i/2) - \cos(\phi_i/2)\right)^2$ does not depend on $\mathtt{S}$. The cost may therefore be minimized by finding the rotation that best aligns the weighted rotation axes, where the axis is weighted (multiplied) by the weight $\sin(\theta_i/2)$ or $\sin(\phi_i/2)$, respectively. Note that in this formulation, the conditions 1 – 3 are still necessary in order to ensure that consistently directed rotation axes are aligned.

Alignment of vectors can be accomplished by the algorithm of Horn [33], which yields an essentially equivalent algorithm to the one already given using quaternions. An alternative method is to use the Procrustes algorithm [21] in which the rotation $\mathtt{S}$ that best aligns vectors $\mathbf{u}_i$ with $\mathtt{S}\mathbf{v}_i$ is the closest rotation matrix (under Frobenius norm) to $\sum_i \mathbf{u}_i {\mathbf{v}_i}^\top$.

Chordal L_2 distance. One could think of trying a similar linear solution to solve the conjugate rotation averaging problem under the chordal L_2-distance as follows. Using the Kronecker product and the vectorization operation, we can rewrite

$$\sum_{i=1}^{n} \|\mathtt{R}_i\mathtt{S} - \mathtt{S}\mathtt{L}_i\|_\mathrm{F}^2 = \sum_{i=1}^{n} \|(\mathtt{R}_i \otimes \mathtt{I} - \mathtt{I} \otimes \mathtt{L}_i^\top)\mathbf{s}\|_2^2 = \|\mathtt{A}\mathbf{s}\|_2^2$$

where $\mathbf{s} = \mathrm{vec}(\mathtt{S}) \in \mathbb{R}^9$ and all the matrices $\mathtt{R}_i \otimes \mathtt{I} - \mathtt{I} \otimes \mathtt{L}_i^\top$ are stacked in one matrix $\mathtt{A} \in \mathbb{R}^{9n\times 9}$. Minimizing this expression could be viewed as a least squares problem and solved through singular value decomposition (SVD). The solution gives a unit length vector $\mathbf{s}$, but one not necessarily corresponding to a rotation matrix. So orthogonal projection onto the special orthogonal group $\mathrm{SO}(3)$ is needed which could be realized using SVD.

This method finds the unit vector $\mathbf{s}$ minimizing $\|\mathtt{A}\mathbf{s}\|_2$, followed by projection onto $\mathrm{SO}(3)$. This is not the same thing as minimizing $\|\mathtt{A}\mathbf{s}\|_2$ directly for $\mathbf{s}$ representing a rotation matrix. Thus the algorithm will not give an optimal result in general.

6.2 Other Closed Form Solutions for Conjugate Averaging

In the robotics community, the following closed-form solution for the hand-eye coordination problem is well known. Park and Martin [64] solved $\mathtt{AX} = \mathtt{XB}$ on the Special Euclidean group, providing a closed-form solution under certain conditions. Here we only treat the case of rotations (no translations), that is $\mathtt{R}_i\mathtt{S} = \mathtt{S}\mathtt{L}_i$ in our notation.

Let $\mathbf{r}_i$ be the angle-axis representation of $\mathtt{R}_i$ that is $\mathbf{r}_i = \log(\mathtt{R}_i)$; correspondingly let $\mathbf{l}_i = \log(\mathtt{L}_i)$. It can be easily verified that $\log(\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}) = \mathtt{S}^{-1}\mathbf{r}_i$. Hence, we have $\|\log(\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}) - \log(\mathtt{L}_i)\|_2 = \|\mathtt{S}^{-1}\mathbf{r}_i - \mathbf{l}_i\|_2$ and we obtain the following objective function:

$$g(\mathtt{S}) = \sum_{i=1}^{n} \|\mathtt{S}^{-1}\mathbf{r}_i - \mathbf{l}_i\|_2^2 = \sum_{i=1}^{n} \|\mathbf{r}_i - \mathtt{S}\mathbf{l}_i\|_2^2 \,.$$

Taking $\mathbf{r}_i$ and $\mathbf{l}_i$ as angle-weighted rotation axes in angle-axis space, minimizing $g(\mathtt{S})$ can be explained as before as searching for the optimal rotation which relates two sets of rotation axes.

Note, however, that the distance measure underlying this idea is the distance $d_{\log}(\mathtt{S},\mathtt{R}) = \|\log(\mathtt{S}) - \log(\mathtt{R})\|_2$ which is not bi-invariant as we have remarked previously. The difference between this solution and the one given above minimizing the L_2 quaternion metric is that the axes $\mathbf{r}_i$ and $\mathbf{l}_i$ are weighted differently. Here, each $\mathbf{r}_i$ or $\mathbf{l}_i$ is weighted by the angle θ_i or ϕ_i of the corresponding rotation. In the quaternion metric case, we weighted by $\sin(\theta_i/2)$ and $\sin(\phi_i/2)$. The resulting solution will be slightly different, because of the different weighting. The previous solution seems more principled, since by adopting the $\sin(\theta_i/2)$, $\sin(\phi_i/2)$ weighting we minimize some meaningful metric. There seems to be no reason to use this solution rather than the quaternion metric solution.

6.3 A Gradient Method for Conjugate Averaging

For the conjugate averaging problem, we can obtain the gradient for the cost function $C(\mathtt{S}) = \sum_{i=1}^{n} d^p(\mathtt{R}_i\mathtt{S}, \mathtt{S}\mathtt{L}_i)$. Thus gradient descent methods can be applied to solve this problem. To compute this gradient from the gradient for the distance measure, the chain rule needs to be applied on $\mathrm{SO}(3)$ rather than in $\mathbb{R}$ as in the previous examples. The details of the gradient computation are given in the appendix. We will only cover the geodesic L_1-mean here and leave the other cases to the interested reader.

Under the angular distance, the cost function is: $C(\mathtt{S}) = \sum_{i=1}^{n} d_\angle(\mathtt{R}_i\mathtt{S}, \mathtt{S}\mathtt{L}_i)$. The gradient for each of the summands $C_{\mathtt{L}_i,\mathtt{R}_i}(\mathtt{S}) = d_\angle(\mathtt{R}_i\mathtt{S}, \mathtt{S}\mathtt{L}_i)$ is

$$\nabla C_{\mathtt{L}_i,\mathtt{R}_i}(\mathtt{S}) = -\mathtt{S}\frac{\log\left(\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}\mathtt{L}_i^\top\right) - \log\left(\mathtt{L}_i^\top\mathtt{S}^{-1}\mathtt{R}_i\mathtt{S}\right)}{d_\angle(\mathtt{R}_i, \mathtt{S}\mathtt{L}_i\mathtt{S}^{-1})}.$$

We propose a Riemannian gradient descent algorithm with geodesic line search to compute the geodesic L_1-mean for the conjugate rotation averaging problem.

```
Choose a tolerance ε > 0.
Set S := d²_quat-mean({L_1, ..., L_n, R_1, ..., R_n}).
loop
    Compute r := Σ_{i=1}^n (log(S^{-1} R_i S L_i^⊤) − log(L_i^⊤ S^{-1} R_i S)) / d_∠(R_i, S L_i S^{-1}).
    Compute s* := argmin_{s≥0} C(S exp(s r)).
    if ||s* r|| < ε then
        return S
    end if
    Update S := S exp(s* r).
end loop
```

Algorithm 4: Computing the geodesic L_1-mean on $\mathrm{SO}(3)$ for the conjugate rotation averaging problem

7 Multiple Rotation Averaging

In this problem, we are given a set of relative rotations, $\mathtt{R}_{ij}$ between coordinate frames indexed by i and j. Only some $\mathtt{R}_{ij}$ are given, represented by index pairs (i, j) in a set $\mathcal{N}$. These relative orientations will in general not be compatible, so the task is to find n rotations $\mathtt{R}_i$ so that $\mathtt{R}_{ij} \approx \mathtt{R}_j\mathtt{R}_i^{-1}$. The appropriate minimization problem is expressed as

$$\underset{\mathtt{R}_1,\ldots,\mathtt{R}_n}{\operatorname{argmin}} \sum_{(i,j)\in\mathcal{N}} d^p(\mathtt{R}_{ij}, \mathtt{R}_j\mathtt{R}_i^{-1})$$

where we are particularly interested in the cases $p = 1$ and $p = 2$ and the above model is to be minimized over all choices of $\mathtt{R}_i, i = 1, 2, \ldots, n$. The distance measures include geodesic, quaternion and chordal.

This problem is a complex multi-variable nonlinear optimization problem. There seems to be no direct method of minimizing this cost function under any of the metrics we consider. In the following subsections, we will first consider two least squares algorithms for quaternion averaging and chordal averaging. Although optimality has been claimed for these algorithms, we show that this will not be the case. We will then discuss the structure of the above cost function in more detail and suggest alternative algorithms.

7.1 Quaternion Averaging for Multiple Rotations

Govindu [22] suggested a method for solving this problem, as follows. Representing the above rotations as quaternions $\mathbf{r}_i$, $\mathbf{r}_j$ and $\mathbf{r}_{ij}$, the equation $\mathtt{R}_{ij}\mathtt{R}_i = \mathtt{R}_j$ can be written in quaternion form as

$$\mathbf{r}_{ij} \cdot \mathbf{r}_i - \mathbf{r}_j = 0\ . \tag{24}$$

Since quaternion multiplication is bilinear in the two operands, this equation gives rise to a set of linear equations in the entries of all the quaternions $\mathbf{r}_i$. The set of all such equations can be written as a set of linear equations of the form $\mathtt{A}\mathbf{r} = \mathbf{0}$, where $\mathbf{r}$ is a vector formed by concatenating all the quaternions $\mathbf{r}_i$. This set of equations may be solved in least-squares enforcing the condition $\|\mathbf{r}\|_2 = \sqrt{n}$.

It has at times been claimed that this algorithm will give a Maximum Likelihood solution under an assumption of Gaussian noise. However, this claim is not valid on at least two counts.

1. Because of the sign ambiguity of the quaternion rotation representation the correct equations should be of the form

   $$\mathbf{r}_{ij} \cdot \mathbf{r}_i - \varepsilon_{ij}\mathbf{r}_j = 0$$

   where $\varepsilon_{ij} = \pm 1$. It is easy to construct examples in which there is no way to assign consistent signs to all the quaternions that will make the equations (24) solvable. A numerical example is given below.
2. Even if the signs of the quaternions can be chosen consistently, then the method does not give the correct mean under any norm, including the quaternion distance. For this to correspond to a true minimum of squared quaternion distance, it would be necessary to minimize $\|\mathtt{A}\mathbf{r}\|_2$ subject to the condition that each of the quaternions $\mathbf{r}_i$ had unit length. Algebraically this can not be done in closed form. Instead, the easy thing is to minimize $\|\mathtt{A}\mathbf{r}\|_2$ subject to the condition that $\mathbf{r}$, the concatenation of all the quaternions, has norm $\sqrt{n}$. In theory, and in practice, this is an entirely different thing from normalizing each of the $\mathbf{r}_i$ separately. Although it generally gives reasonable results, it certainly does not give the optimal result under any sensible distance.

7.1.1 Problem statement

The basic formulation of the multiple rotation averaging problem in quaternion representation is

$$\mathbf{r}_{ij}\mathbf{r}_i - \varepsilon_{ij}\mathbf{r}_j = 0 \tag{25}$$

where $\varepsilon_{ij} = \pm 1$. The quaternions $\mathbf{r}_{ij}$ representing the relative rotations are supposed known, and the task is to find the quaternions $\mathbf{r}_i$, $\mathbf{r}_j$ that satisfy this equation, for a set of given pairs (i, j). We will look at ways of determining the signs ε_{ij} which will make these equations true, and hence will allow us to find a solution.

First, we will see how these equations look, when written in terms of matrices. We define a matrix $\mathtt{R}_{ij}^{\times}$ that corresponds to the quaternion multiplication. Let $\mathbf{r}_{ij}$ be written as a quaternion $(c, \mathbf{v})$ where $c = \cos(\theta/2)$ and θ is the rotation angle; $\mathbf{v}$ is a vector of length $\sin(\theta/2)$ representating the rotation axis. Since $\theta \le \pi$, we may choose $c \ge 0$. Then multiplication of a quaternion $\mathbf{r}_i$ by $\mathbf{r}_{ij}$ is equivalent to the matrix product

$$\begin{aligned}\mathbf{r}_j &= \mathtt{R}_{ij}^{\times}\,\mathbf{r}_i \\ &= \begin{bmatrix} c & -\mathbf{v}^\top \\ \mathbf{v} & [\mathbf{v}]_\times + c\,\mathtt{I}_{3\times 3}\end{bmatrix}\mathbf{r}_i\end{aligned} \tag{26}$$

Lemma 5 *The matrix appearing in (26) is orthogonal, meaning that* $\mathtt{R}_{ij}^{\times}\,\mathtt{R}_{ij}^{\times\top} = \mathtt{I}_{4\times 4}$. *Furthermore, for any vector* $\mathbf{r}_i$, *we have* $\mathbf{r}_i{}^\top\,\mathtt{R}_{ij}^{\times}\,\mathbf{r}_i \ge 0$. *Consequently, the angle between* $\mathbf{r}_i$ *and* $\mathtt{R}_{ij}^{\times}\,\mathbf{r}_i$ *is no greater than* $\pi/2$.

To show this, observe that

$$\mathbf{r}_i{}^\top \begin{bmatrix} c & -\mathbf{v}^\top \\ \mathbf{v} & [\mathbf{v}]_\times + c\mathtt{I}\end{bmatrix}\mathbf{r}_i \;=\; \mathbf{r}_i{}^\top\, c\mathtt{I}_{4\times 4}\,\mathbf{r}_i \;=\; c \;\ge\; 0,$$

since the skew-symmetric parts of the matrix do not contribute to the product.

Using this representation of quaternion multiplication, the set of equations (25) forms a $4m \times 4n$ set of equations, where n is the number of rotations, and m is the number of pairs (i, j). This set of equations can be written as $\mathtt{M}\mathbf{r} = \mathbf{0}$, where $\mathbf{r}$ is a vector made up by concatenating the components of all the quaternions. In the presence of noise, we find the least-squares solution using Singular Value Decomposition (SVD). Writing $\mathtt{M} = \mathtt{UDV}^\top$, the required solution for $\mathbf{r}$ is the last column of $\mathtt{V}$. To obtain unit quaternions that represent rotations, we need to normalize each of the $\mathbf{r}_i$ individually to unit length. Here $\mathbf{r}_i$ represents the 4-vector containing the block of 4 entries in $\mathbf{r}$ corresponding to the i-th rotation.

Example. We illustrate the need for the signs ε_{ij} with a specific example. Consider three rotations $\mathtt{R}_1$, $\mathtt{R}_2$ and $\mathtt{R}_3$ and measured relative rotations $\mathtt{R}_{12} = \mathtt{R}_{23} = \mathtt{R}_{31}$, each being a rotation through $120°$ about the x axis. Obviously, this represents a coordinate frame undergoing one complete rotation.

The quaternion corresponding to the relative rotation is $(1/2, \sqrt{3}/2, 0, 0)$ and left-multiplication by this quaternion is represented by the 4×4 matrix

$$\mathtt{M} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 & 0 & 0 \\ \sqrt{3}/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & -\sqrt{3}/2 \\ 0 & 0 & \sqrt{3}/2 & 1/2 \end{bmatrix}$$

The complete set of equations (24) may be written as a matrix equation

$$\begin{bmatrix} \mathtt{M} & -\mathtt{I} & \mathtt{0} \\ \mathtt{0} & \mathtt{M} & -\mathtt{I} \\ -\mathtt{I} & \mathtt{0} & \mathtt{M} \end{bmatrix} \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{pmatrix} = \mathbf{0}. \tag{27}$$

It is easily verified that this matrix has determinant 16, so there is no exact solution to the set of equations.

7.1.2 Algorithm statement

The complete algorithm is given as follows.

1. Given relative rotations $\mathtt{R}_{ij}$, choose a quaternion representation $\mathbf{r}_{ij}$ for each.
2. Find coefficients $\varepsilon_{ij} = \pm 1$ such that (25) will hold for the true solution.
3. Form a set of matrix equations $\mathtt{M}\mathbf{r} = \mathbf{0}$ using (26) and take the SVD, $\mathtt{M} = \mathtt{UDV}^\top$. The solution is a vector $\mathbf{r} = ({\mathbf{r}_1}^\top, \ldots, {\mathbf{r}_n}^\top)^\top$, namely the last column of $\mathtt{V}$.
4. Normalize each $\mathbf{r}_i$ to $\mathbf{r}_i / \|\mathbf{r}_i\|_2$ to give a solution for each of the equations.

Previous versions of this algorithm have ignored the need to select the correct signs here, and have therefore solved the wrong equations. Without the correct signs ε_{ij}, the equations (25) may not have a solution as the example above shows. The signs ε_{ij} may be chosen using the following simple algorithm.

1. Choose all the relative rotation quaternions $\mathbf{r}_{ij}$ so that the first coefficient (real part of the quaternion) is non-negative.
2. Select a tree in the graph formed by joining nodes corresponding to the $\mathbf{r}_i$ with an edge, when $\mathbf{r}_{ij}$ is defined.
3. Assign an initial value $\mathbf{r}_i = 0$ to some node chosen as the root of the tree, and propagate the estimate of $\mathbf{r}_j$ across the tree using the relations $\mathbf{r}_j = \mathbf{r}_{ij}\mathbf{r}_i$ and set $\varepsilon_{ij} = +1$ for an edge in the tree.
4. For an edge $\mathbf{r}_{ij}$ not in the tree, set $\varepsilon_{ij} = +1$ or -1 depending on whether $\mathbf{r}_{ij}\mathbf{r}_i$ is closest to $\mathbf{r}_j$ or $-\mathbf{r}_j$.

Unless there is a large accumulated error in the rotations as they are propagated over the tree, the decision of which value of ε_{ij} to choose should be clear.

Note that in solving these equations, we find the solution such that $\|\mathbf{r}\|_2 = 1$. To be more correct, we should minimize the cost $\|\mathtt{M}\mathbf{r}\|_2$ subject to the constraint that each individual $\mathbf{r}_i$ has unit norm. However, this is not possible by linear means, and is probably a hard problem in general. If we could solve subject to these constraints, then the solution would be the true least-squares solution minimizing squared quaternion distance (the distance metric measuring distance between quaternion representations of rotations).

7.2 Chordal Averaging for Multiple Rotations

Chordal L_2-averaging for the multiple rotation averaging problem is described as finding the rotations minimizing the cost

$$\sum_{(i,j)\in\mathcal{N}} \|\mathtt{R}_{ij}\mathtt{R}_i - \mathtt{R}_j\|_{\mathrm{F}}^2 \,.$$

Without enforcing the orthogonality constraints, we can solve the above model as a least squares problem through vectorization and singular value decomposition. Finally, all the orthogonal constraints are enforced through subsequently finding the nearest orthogonal matrices by polar decomposition [55]. According to the analysis in [55], the chordal averaging algorithm performs better than the quaternion averaging algorithm due to the availability of 9 parameters for each rotation instead of only 4 in the quaternion representation. However, the solution will in general not be optimal.

Unlike the quaternion method, the method involving matrices does not suffer from the issue of needing to select the correct sign for the quaternion.

7.3 The Structure of the Cost Function for Multiple Rotations

In this section, we will take a closer look at the cost function

$$C(\mathtt{R}_1, \ldots, \mathtt{R}_n) = \sum_{(i,j)\in\mathcal{N}} d(\mathtt{R}_{ij}, \mathtt{R}_j\mathtt{R}_i^{-1})^p \tag{28}$$

for the multiple rotation averaging problem. The question we will consider is the convexity of this cost function as a function of the rotations $\mathtt{R}_i$. The results we obtain will be largely negative, particularly for the L_2 cost functions ($p = 2$). We will exhibit examples where the residual cost is arbitrarily small, at a local minimum, but the global minimum lies in a different basin of attraction. Furthermore, it can be shown that this cost function usually has saddle points.

One of the results we obtained (theorem 5) for several of the distance measures in the single rotation averaging problem (estimate $\mathtt{R}$ given rotation estimates $\mathtt{R}_i$) was that if all the $\mathtt{R}_i$ lie in a convex set (for instance, an open ball of radius $\pi/2$), then the optimal solution lies in the convex set and the cost function is convex on this set. Thus, once we have found an estimate $\mathtt{R}$ with sufficiently small residual (less than $\pi/2$) for each $\mathtt{R}_i$, the optimum can be found by convex optimization techniques. It will be shown that this is not the case in the multiple rotation estimation problem.

An example. We give an example based on the intuition that if a vehicle with an inertial rotation sensor follows a long closed path, returning to its initial position, then it may be difficult to determine whether the vehicle has rotated through a complete turn or not during the trajectory. Thus, consider the case where we wish to estimate rotations $\mathtt{R}_i; i = 0, \ldots, n-1$ when estimates $\mathtt{R}_{ij}$ are known only for consecutive positions ($j = i+1$), as well as for the initial and final positions $R_{n-1,0}$.

Suppose that all rotations are about a single (perhaps vertical) axis, and that in the true solution, $\mathtt{R}_i$ is a rotation through an angle $2\pi i/n$. Suppose that the relative rotations are measured accurately, so that $\angle(\mathtt{R}_{i,i+1}) = \angle(\mathtt{R}_{n-1,0}) = 2\pi/n$. Clearly in this case, $\mathtt{R}_{i,i+1} = \mathtt{R}_{i+1}\mathtt{R}_i{}^\top$ exactly, for $i = 0, \ldots, n-1$,[2] so that the true solution has zero cost. However, there is a different solution that may have small cost, namely $\mathtt{R}_i = \mathtt{I}$ for all i. For instance in the squared angular distance case, the cost will be

$$\begin{aligned} C &= \sum_{i=0}^{n-1} d_\angle(\mathtt{R}_{i,i+1}, \mathtt{R}_{i+1}\mathtt{R}_i{}^\top)^2 = \sum_{i=0}^{n-1} d_\angle(\mathtt{R}_{i,i+1}, \mathtt{I})^2 \\ &= \sum_{i=0}^{n-1} (2\pi/n)^2 = 4\pi^2/n \end{aligned}$$

which can be arbitrarily small for large n.

For a slightly different example, if each of the measured angles is $\angle(\mathtt{R}_{i,i+1}) = \pi/n$, then the two solutions will have equal cost $\sum_{i=1}^n (\pi/n)^2 = \pi^2/n$ which can also be made arbitrarily small by choosing n large.

Basins of attraction. It may be thought that in the first example given here, with $\mathtt{R}_i = \mathtt{I}$ that this solution may be continuously modified to the minimum solution given by $\angle(\mathtt{R}_i) = 2\pi i/n$. However, it will be shown that this is not the case. In fact, these two solutions lie in different basins of attraction in the cost "surface".

Most continuous optimization techniques act by modifying a current solution by iteratively moving from one potential solution to another, usually in a direction of decreasing cost. Although the sequence of iterates is finite, the process may be approximated by the estimate traversing a continuous path across the cost surface from an initial solution to a final solution. If a continuous downhill path exists to a minimum, then the likelihood of reaching this minimum is much higher. Given a local minimum of a cost function, one may define its *basin of attraction* to be the set of points that are connected to the given local minimum by a decreasing cost path.

It will be shown that the two solutions in the example given above lie in different basins of attraction, and hence one can not go from one to the other by a downhill path.

Consider an n-tuple of rotations $(\mathtt{R}_0, \ldots, \mathtt{R}_{n-1}) \in \mathrm{SO}(3)^n$ where $\mathtt{R}_0 = \mathtt{I}$; we define also $\mathtt{R}_n = \mathtt{R}_0 = \mathtt{I}$. We think of this n-tuple as being an estimate of the solution to an n-rotation averaging problem defined by a set of relative rotations $\mathtt{R}_{i,i+1}$. The cost function defines a function from $\mathrm{SO}(3)^n$ to $\mathbb{R}$, defining a cost for such an n-tuple of measurements. We suppose that there is a continuous family of such n-tuples, $(\mathtt{R}_0^t, \ldots, \mathtt{R}_{n-1}^t)$ for $t \in [0,1]$, tracing out a path in $\mathrm{SO}(3)^n$, transforming an initial estimate $(\mathtt{R}_0^0, \ldots, \mathtt{R}_{n-1}^0)$ to a final estimate $(\mathtt{R}_0^1, \ldots, \mathtt{R}_{n-1}^1)$. We also set $\mathtt{R}_n^t = \mathtt{R}_0^t = \mathtt{I}$ for all t.

Now, we focus on an n-tuple defined for a given fixed value of t, and use it to define a closed path in $\mathrm{SO}(3)$, based at the identity rotation $\mathtt{R}_0^t$. The idea is to think of these n rotations as being sampled positions from a continuously varying coordinate frame traversing a closed path in rotation space, $\mathrm{SO}(3)$. The continuous path is obtained by filling in between the rotations $\mathtt{R}_i^t, \mathtt{R}_{i+1}^t$ by interpolation along the shortest geodesic. The resulting path in rotation space may be intuitively thought of as the estimate (at parameter value t) of the path of the coordinate frame through rotation space.

More formally, for a fixed t, we use the n-tuple $(\mathtt{R}_0^t, \ldots, \mathtt{R}_{n-1}^t)$ to define a closed path $\gamma_t(s)$ in $\mathrm{SO}(3)$. This path is defined as follows. Define $\mathtt{R}_n^t = \mathtt{R}_0^t$. Now, for each $s \in [0,1]$ we wish to define a point (rotation) in $\mathrm{SO}(3)$. For $s = i/n$ for some $i = 0, \ldots, n$, we define $\gamma_t(s) = \mathtt{R}_i^t$.

[2] For convenience of notation, we consider the index n to mean 0, so that $\mathtt{R}_{i+1}$ means $\mathtt{R}_0$ and $\mathtt{R}_{i,i+1}$ means $\mathtt{R}_{n-1,0}$ when $i = n-1$.

This defines the path γ_t at evenly spaced point $s \in [0,1]$. We wish to interpolate this path to all values of s. This is done by interpolating along geodesics. Thus, suppose that $i/n < s < (i+1)/n$ for some i. Then, for s in the interval $[i/n, (i+1)/n]$, the path $\gamma_t(s)$ moves with constant velocity along the shortest geodesic from $\mathtt{R}_i$ to $\mathtt{R}_{i+1}$. Thus, for each t, the path $\gamma_t(s)$ is a continuous path in SO(3). Since $\mathtt{R}_0^t = \mathtt{R}_n^t$, we see that $\gamma_t(s)$ is a closed path based at $\mathtt{R}_0^t$. Note that at time t, the path $\gamma_t(s)$ so defined corresponds intuitively to the current estimate (at time t) of the path of the coordinate frame in SO(3).

Since each of the rotations $\mathtt{R}_i^t$ traces out continuous paths in SO(3) as t varies, we may define a mapping $\gamma : [0,1] \times [0,1] \rightarrow \mathrm{SO}(3)$ as $\gamma(t,s) = \gamma_t(s)$. Our purpose is to show the following properties:

1. γ is a continuous mapping.
2. For each t, $\gamma(t,0) = \gamma(t,1) = \mathtt{I}$.

Under these circumstances, we say that the two paths $\gamma_0(s)$ and $\gamma_1(s)$ are *homotopic* or *homotopy equivalent* as closed paths based (starting and ending) at the base point $\mathtt{I} \in \mathrm{SO}(3)$. Under the equivalence relationship of homotopy, based paths in SO(3) form the *fundamental group* $\pi_1(\mathrm{SO}(3), \mathtt{I})$. Under the two conditions given above, the two paths $\gamma_0(s)$ and $\gamma_1(s)$ represent the same element of the fundamental group.

It is well known that the fundamental group of SO(3) is equal to Z_2, the group with two elements. This is easily seen, since the mapping from the geodesic sphere S^3 to SO(3) is a 2-fold covering, and $\pi_1(S^3)$ is the trivial group with one element. (Using another common terminology, S^3 is simply-connected.)

We now look at the first of the two conditions given above, namely that γ should be a continous mapping. Consider a point (t,s) with $i/n \leq s \leq (i+1)/n$. Then the point $\gamma(t,s)$ lies on the shortest geodesic from $\mathtt{R}_i^t$ to $\mathtt{R}_{i+1}^t$. As t varies, the rotations $\mathtt{R}_i^t$ and $\mathtt{R}_{i+1}^t$ vary continuously. If the shortest geodesic between these two rotations also varies continuously, then $\gamma(t,s)$ will move as a continuous function of t and s. There are in general two geodesic paths between any two points (rotations) in SO(3), corresponding to different arcs of the great circle in the quaternion sphere. However, if the angular distance between $\mathtt{R}_i^t$ and $\mathtt{R}_{i+1}^t$ remains less than π, then the shorter of the two geodesics will be unambiguously defined, and the geodesic will move continuously with its end points. Thus, we have shown the following result.

Lemma 6 *Let* $(\mathtt{R}_0^t, \ldots, \mathtt{R}_n^t)$ *with* $\mathtt{R}_0^t = \mathtt{R}_n^t = \mathtt{I}$ *be rotation estimates continuously varying for* $t \in [0,1]$*, from an initial estimate when* $t = 0$ *to a final estimate when* $t = 1$*. Define paths* $\gamma_t(s)$ *in* SO(3) *by the construction above, interpolating between the rotations* $\mathtt{R}_i^t$ *for fixed values of* $t \in [0,1]$*. Suppose that* $d_\angle(\mathtt{R}_i^t, \mathtt{R}_{i+1}^t) < \pi$ *for all* t *and all* i*. Then the paths* $\gamma_0(s)$ *and* $\gamma_1(s)$ *are homotopy equivalent.*

From this we may deduce that if the two paths γ_0 and γ_1 are not homotopy equivalent, then at some point t between 0 and 1, and for some value of j, $d_\angle(\mathtt{R}_j^t, \mathtt{R}_{j+1}^t) = \pi$. This means that the cost of the intermediate solution $(\mathtt{R}_1^t, \ldots, \mathtt{R}_n^t)$ must satisfy

$$
\begin{aligned}
C(\mathtt{R}_0^t, \ldots, \mathtt{R}_{n-1}^t) &= \sum_{i=0}^{n-1} d_\angle(\mathtt{R}_{i,i+1}^t, \mathtt{R}_{i+1}^t \mathtt{R}_i^{t\top})^2 \\
&\geq d_\angle(\mathtt{R}_{j,j+1}^t, \mathtt{R}_{j+1}^t \mathtt{R}_j^{t\top})^2 \\
&\geq (d_\angle(\mathtt{R}_j^t, \mathtt{R}_{j+1}^t) - d_\angle(\mathtt{R}_{j,j+1}^t, \mathtt{I}))^2 \\
&\geq (\pi - d_\angle(\mathtt{R}_{j,j+1}^t, \mathtt{I}))^2 \ ,
\end{aligned}
$$

where the second-last line follows from the triangle inequality.

Therefore, to transform an initial estimate of the rotations to a final estimate, where the initial and final interpolated trajectories $\gamma_0(s)$ and $\gamma_1(s)$ are not homotopy equivalent, must involve an intermediate estimate which has cost greater than the above value. If the initial and final estimates have smaller cost than this, then they must lie in different basins of attraction and to get from one to the other must require an intermediate estimate of large cost.

Finally, we show that in the example we gave above, the two paths γ_0 and γ_1 are not homotopy equivalent, since one path contains a rotation through 2π and the other one does not.

In one case, $\mathtt{R}_i^1 = \mathtt{I}$, the interpolated path is $\gamma_1(s) = \mathtt{I}$, that is, the path is constant at the base point $\mathtt{I}$. In the true solution, all the rotations are about the same axis, and $\angle(\mathtt{R}_i^0) = 2\pi i/n$. From this we see that the interpolated path is given by $\gamma_0(s) = \mathtt{R}_s^0$ with $\angle(\mathtt{R}_s) = 2\pi s$. During this path the rotation turns through one complete turn through 2π radians about the rotation axis. However, this is not a null-homotopic path in SO(3), since when lifted to the 2-fold covering space, namely the quaternion sphere, it lifts to a path from the quaternion $\mathbf{r} = (1,0,0,0)$ to $(-1,0,0,0)$.

We can conclude that to pass from the wrong solution $\mathtt{R}_i^1 = \mathtt{I}$ with cost $4\pi^2/n$ to the correct solution $\angle(\mathtt{R}_i^0) = 2\pi i/n$ with zero cost, a continuous optimization scheme would have to overcome a hurdle of cost at least $(\pi - 2\pi/n)^2$, which is much larger than the cost of the wrong solution, $4\pi^2/n$, for large n.

7.4 An Iterative Algorithm for Multiple Rotation Averaging

As discussed in the previous sections, there seems to be no direct method of minimizing the multiple rotation averaging cost function under any of the distances we consider. Therefore, our strategy is to minimize the cost function by using rotation averaging to update each $\mathtt{R}_i$ in turn. At each step of this algorithm, the total cost decreases, and hence *the cost* converges to a limit. We do not at present claim a rigorous proof that the algorithm converges to even a local minimum. We do know that the sequence of estimates must contain a convergent subsequence, and the limit of this subsequence must be at least a local minimum with respect to each $\mathtt{R}_i$ individually. In light of the existence of saddle points in the cost function this is however a relatively weak result.

Initial values for each $\mathtt{R}_i$ are easily found by propagating from a given rotation $\mathtt{R}_0$ assumed to be the identity.

```
1: Set t:=0 and pick initial values R_1^(0), ..., R_n^(0).
2: loop
3:    for j=1,...,n do
4:       Set R_j^(t+1) := d^p-mean({R_ij R_i^(t)}_(i,j)∈N).
5:    end for
6:    t ← t + 1.
7: end loop
```

Algorithm 5: An iterative algorithm for multiple rotation averaging

We term Algorithm 5 a block Jacobi type algorithm because Step 4 entails a minimization of C over the jth factor in $\mathrm{SO}(3)^n$ while the other variables are being kept constant. Steps 3-5 hence contain a Jacobi sweep over the full parameter space. Since $\mathrm{SO}(3)$ is 3-dimensional, this corresponds to a block version of a classical Jacobi type algorithm where each inner minimization would be carried out over a 1-dimensional curve.

The convergence of block Jacobi type methods on manifolds has been studied by Hüper [36], but at this stage we haven't been able to successfully apply this theory to the particular cost function at hand.

7.5 L_1 averaging multiple rotations

The iterative averaging scheme described in the previous section may be used for L_1 geodesic multiple rotation averaging by using successive applications of the Weiszfeld algorithm. At any given point during the computation, a rotation $\mathtt{R}_j$ will have an estimated value, and so will its neighbours $\mathtt{R}_i$, for $(i, j) \in \mathcal{N}$. Therefore, we may compute estimates $\mathtt{R}_j^{(i)} = \mathtt{R}_{ij}\mathtt{R}_i$, where the superscript (i) indicates that this is the estimate of $\mathtt{R}_j$ derived from its neighbour $\mathtt{R}_i$. We then use our Weiszfeld L_1 averaging method on $\mathrm{SO}(3)$ to compute a new estimate for $\mathtt{R}_j$ by averaging the estimates $\mathtt{R}_j^{(i)}$. In one pass of the algorithm, each $\mathtt{R}_j$ is re-estimated in turn, in some order. Multiple passes of the algorithm are required for convergence.

Since the Weiszfeld algorithm on $\mathrm{SO}(3)$ is itself an iterative algorithm, we have the choice of running the Weiszfeld algorithm to convergence, each time we re-estimate $\mathtt{R}_j$, or else running it for a limited number of iterations leaving the convergence incomplete, and passing on to the next rotations. To avoid nested interation, we choose to run a single iteration of the Weiszfeld algorithm at each step. The complete algorithm is as follows.

1. **Initialization:** Set some node $\mathtt{R}_{i_0}$ with the maximum number of neighbours to the identity rotation, and construct a spanning tree in the neighbourhood graph rooted at $\mathtt{R}_{i_0}$. Estimate the rotations $\mathtt{R}_j$ at each other node in the tree by propagating away from the root using the relation $\mathtt{R}_j = \mathtt{R}_{ij}\mathtt{R}_i$.
2. **Sweep:** For each j in turn, re-estimate the rotation $\mathtt{R}_j$ using one iteration of the Weiszfeld algorithm. (As each new $\mathtt{R}_j$ is computed, it is used in the computations of the other $\mathtt{R}_j$ during the same sweep.)
3. **Iterate:** Repeat this last step a fixed number of times, or until convergence.

The whole computation is most conveniently carried out using quaternions.

Unlike the single rotation averaging problem considered in section 5 we can not guarantee convergence of this algorithm to a global minimum, but initial simulation results demonstrate good performance, see [27].

7.6 Summary for Multiple Rotation Averaging

For the multiple rotation averaging problem, there seems to be no direct optimization method on $\mathrm{SO}(3)^n$. We have shown that the associated cost function usually exhibits nontrivial structure, including saddle points and multiple local minima in separate basins of attraction. We propose two algorithms: iterative averaging, and Weiszfeld based L_1 averaging.

Acknowledgements This work was partially supported by NICTA, a research laboratory funded by the Australian Government, in part through the Australian Research Council.

Appendix – Convexity

Of major relevance to questions of convergence and uniqueness of solutions of averaging problems is determining if and where the defined cost functions are convex functions.

In this section we consider the question of convexity of a function measuring distance in $\mathrm{SO}(3)$ from a given rotation $\mathtt{R}$. Since we are dealing with a function defined on $\mathrm{SO}(3)$, rather than a Euclidean space, we will need the concept of geodesic convexity to analyze this problem.

The general definition of convexity of a function in $\mathbb{R}^n$ is as follows. Given a convex region $U \subset \mathbb{R}^n$ a function f defined on U is convex if for any two points $\mathbf{x}_0$ and $\mathbf{x}_1$ in U, and any point $\mathbf{y}$ lying on the line segment bounded by $\mathbf{x}_0$

and $\mathbf{x}_1$, given by $\mathbf{y} = (1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x}_1$ with $0 \leq \lambda \leq 1$, we have

$$f(\mathbf{y}) \leq (1-\lambda) f(\mathbf{x}_0) + \lambda f(\mathbf{x}_1) \ .$$

In adapting this definition to $\mathrm{SO}(3)$, or indeed to any Riemannian or differentiable manifold, the role of a line is naturally taken by a geodesic. The appropriate definition of a convex set in $\mathrm{SO}(3)$ is a little less clear, and will be considered next.

Convex sets in $\mathrm{SO}(3)$

As discussed in section 4 the geodesics on $\mathrm{SO}(3)$ are doubly covered by great circles on S^3 and there is a uniform length scaling by a factor of 2 between the geodesics on $\mathrm{SO}(3)$ and those on S^3. In particular, we see that the geodesics on $\mathrm{SO}(3)$ are closed curves with a total length of 2π. There are exactly two geodesic segments between any two points in $\mathrm{SO}(3)$ (without exception). Given two points (rotations) $\mathtt{R}_0$ and $\mathtt{R}_1$ in $\mathrm{SO}(3)$, we call the shorter of the two geodesic segments from $\mathtt{R}_0$ to $\mathtt{R}_1$ the *short geodesic segment* between these points. If $\mathtt{R}_0$ and $\mathtt{R}_1$ differ by a rotation through π, then which of the two geodesic segments is the shorter one is ambiguous and hence there is no short geodesic segment between such points.

We can now define two slightly different notions of geodesic convexity of sets in $\mathrm{SO}(3)$. (The definition is generalizable to other manifolds.)

Definition 1 A non-empty region $U \subset \mathrm{SO}(3)$ is called *weakly convex* if for any two points $\mathtt{R}_0$ and $\mathtt{R}_1$ in U *exactly one* geodesic segment from $\mathtt{R}_0$ to $\mathtt{R}_1$ lies entirely inside U.

A weakly convex region $U \subset \mathrm{SO}(3)$ is called *convex* if the geodesic segment from $\mathtt{R}_0$ to $\mathtt{R}_1$ in U is always the short geodesic segment between these points, having length strictly smaller than π.

The empty set is not considered to be convex or weakly convex.

A closed ball of radius $r \geq 0$ in $\mathrm{SO}(3)$ is a set

$$B(\mathtt{R}, r) = \{\mathtt{S} \in \mathrm{SO}(3) \,|\, d_\angle(\mathtt{S}, \mathtt{R}) \leq r\}$$

for some $\mathtt{R}$ in $\mathrm{SO}(3)$.

Radius and Diameter. We introduce two useful pieces of terminology, the radius and diameter of a set. The diameter of a set C in $\mathrm{SO}(3)$ is the supremum of $d_\angle(\mathtt{R}, \mathtt{S})$ over all $\mathtt{R}, \mathtt{S} \in C$. According to this definition, the diameter of a convex set is at most equal to π, moreover, no two points in the set actually achieve this bound.

An open ball of radius $r > 0$ in $\mathrm{SO}(3)$, denoted $\mathring{B}(\mathtt{R}, r)$, is the interior of the closed ball, consisting of rotations at distance strictly less than r from $\mathtt{R}$. We emphasize for clarity that the balls $B(\mathtt{R}, r)$ or $\mathring{B}(\mathtt{R}, r)$ are defined in terms of the *geodesic* (angular) distance on $\mathrm{SO}(3)$.

The radius of a set C in $\mathrm{SO}(3)$ is the infimum of all r such that C is contained in some ball of radius r. It is evident by the triangle inequality that radius is at least half the diameter of the set.

Lemma 7 *A closed ball in* $\mathrm{SO}(3)$ *is convex if and only if its radius is less than* $\pi/2$. *Similarly, an open ball in* $\mathrm{SO}(3)$ *is convex if and only if its radius is less than or equal to* $\pi/2$. *A closed ball in* $\mathrm{SO}(3)$ *is weakly convex if and only if its radius is less than* π, *and an open ball in* $\mathrm{SO}(3)$ *is weakly convex if and only if its radius is less than or equal to* π.

If we visualize this in terms of the quaternion sphere, the proof is straightforward, and hence omitted. Note that an open ball of radius π is the whole of $\mathrm{SO}(3)$ except for one plane, consisting of rotations at distance π from the centre of the ball.

Convex and weakly convex subsets of $\mathrm{SO}(3)$ can not be arbitrarily "large", in the following precise sense.

Theorem 10 *Any weakly convex subset of* $\mathrm{SO}(3)$ *is contained in an open ball of radius* π. *In other words, there exists a plane in* $\mathrm{SO}(3)$ *(the boundary of the open ball) that does not meet the said weakly convex set. Any convex subset of* $\mathrm{SO}(3)$ *is contained in a closed ball of radius* $2\pi/3$.

The proof of this theorem turns out to be surprisingly difficult (particularly the first part) and will be reported elsewhere [30]. As a consequence of this result we may picture any weakly convex subset of $\mathrm{SO}(3)$ simply as a convex set in $\mathbb{R}^3$ under a suitably chosen gnomonic projection, namely the one mapping the boundary of the containing ball of radius π to the plane at infinity (cf. section 3.4). This is because the gnomonic projection maps geodesics to geodesics, and hence weakly convex sets to convex sets.

Although we provide no proof here, we nevertheless make frequent use of the result of theorem 10 for weakly convex sets. However, in a sense the rest of the paper does not depend on this result, as long as we are willing to modify the definition of weakly convex set to include the (redundant) condition that such a set lies inside an open ball of radius π.

According to this theorem, the radius of a convex set is at most $2\pi/3$, and a closed convex set must have radius strictly less than $2\pi/3$. On the other hand, lemma 7 states that a convex **ball** can have radius no greater than $\pi/2$. It is therefore somewhat surprising that we claim that a ball of radius $2\pi/3$ is required to contain any convex set. This bound is tight however, as a simple example shows. Consider a regular tetrahedron in $\mathbb{R}^3$, centred at the origin. The inverse gnomonic map will take this to a tetrahedron in $\mathrm{SO}(3)$ bounded by geodesic planes. Let the size of this tetrahedron be such that its vertices are at geodesic distance

$2\pi/3$ from its centre. Knowing that the angle α between the vectors from the origin to any two vertices of a regular tetrahedron is given by $\cos(\alpha) = -1/3$, it may be verified directly using (13) (the cosine rule) that the angular distance between two vertices of the tetrahedron is equal to π. It follows from this that for each vertex A of the tetrahedron, the whole geodesic plane passing through the three other vertices lies at distance π from A. Consequently, no two points in the tetrahedron lie at a greater distance than π from each other. The interior of the tetrahedron is therefore convex, contained in a closed ball of radius $2\pi/3$, but not in any closed ball of lesser radius.

Observe that we may add a single vertex (or even the whole boundary, less one face) to this tetrahedron and it will still be convex, but will not lie in an open ball of radius $2\pi/3$; thus we cannot replace the words "closed ball" with "open ball" in the theorem statement. Furthermore, the complete closed tetrahedron (although weakly convex) is not convex, since it contains points at an angular distance π from each other.

Some results about weakly convex sets in $\mathrm{SO}(3)$ follow easily from corresponding statements about convex sets in $\mathbb{R}^3$.

Proposition 3 *Let B be a set in* $\mathrm{SO}(3)$.

1. *If B is a weakly convex set of radius $r < \pi$, then the closure of B is weakly convex.*
2. *If B is a convex set of diameter $d < \pi$, then the closure of B is convex.*
3. *If B is a closed or open weakly convex set, then for any point $\mathbf{x} \notin B$, there exists a plane through $\mathbf{x}$ that does not intersect B.*
4. *If B is a closed or open weakly convex set, then $B = \mathrm{SO}(3) \setminus \bigcup \Pi_i$, where Π_i runs over all planes not intersecting B.*

Proof. We select a plane not containing B and map it to the plane at infinity. The set B is thereby mapped to a convex set in $\mathbb{R}^3$. In the case when B has radius $r < \pi$, this mapping can be chosen so that B maps to a bounded set. The four parts of the theorem then all follow from properties of convex sets in $\mathbb{R}^n$. The corresponding properties of sets in $\mathbb{R}^n$ are not quite trivial. The reader is referred to [69] for the required proofs. □

Separation properties of convex sets by planes are important in the study of convex sets in $\mathbb{R}^n$. The basic separability property in $\mathbb{R}^n$ is that two disjoint convex open sets are separable by a plane ([69], theorem 11.3). As the following results show, similar properties hold for weakly convex sets in $\mathrm{SO}(3)$, but this does not follow immediately from the $\mathbb{R}^n$ case. The necessary modification reflects the fact that a single plane in $\mathrm{SO}(3)$ does not separate $\mathrm{SO}(3)$ into two parts (but *two* planes do).

Proposition 4 *If S and T are two disjoint open weakly convex sets in* $\mathrm{SO}(3)$*, then there exists a plane Π that intersects neither of them.*

Proof. Consider a plane disjoint from S, and identify it as Π_∞, the plane at infinity. If Π_∞ is disjoint from T, then it is the required plane. Otherwise, T is cut into two parts by Π_∞, such that $T_1 \cup T_2 = T \setminus \Pi_\infty$, and T_1 and T_2 are open convex sets in $\mathbb{R}^3$. We form the set $S' = \bigcup L(\mathbf{x}, \mathbf{y})$ where $L(\mathbf{x}, \mathbf{y})$ is a line segment in $\mathbb{R}^3$ joining a point $\mathbf{x} \in S$ and a point $\mathbf{y} \in T_1$, and S' is the union of all such line segments. We claim that S' is the convex hull (in $\mathbb{R}^3$) of $S \cup T_1$.

To see this, consider two points $\mathbf{a}$ and $\mathbf{b}$ in S', where $\mathbf{a}$ is on a line $L(\mathbf{x}_1, \mathbf{y}_1)$ and $\mathbf{b}$ is on a line $L(\mathbf{x}_2, \mathbf{y}_2)$. Now, the points $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{y}_1$ and $\mathbf{y}_2$ are the vertices of a tetrahedron. (The case where the four points are coplanar is a special case which is easily treated separately.) This tetrahedron is convex, and hence contains the line segment from $\mathbf{a}$ to $\mathbf{b}$. Furthermore, every point in the tetrahedron lies on some line with endpoints in the line segments $\mathbf{x}_1\mathbf{x}_2$ and $\mathbf{y}_1\mathbf{y}_2$, which lie inside S and T_1 respectively. Hence the whole tetrahedron, and in particular the line segment from $\mathbf{a}$ to $\mathbf{b}$, lies inside S'.

Now, we claim that this convex set S' is disjoint from T_2. In particular, if a point $\mathbf{a} \in T_2$ lies on the line segment $L(\mathbf{x}, \mathbf{y})$, with $\mathbf{x} \in S$, $\mathbf{y} \in T_1$, then both $\mathbf{a}$ and $\mathbf{y}$ lie in T, which is by assumption weakly convex. A line segment from $\mathbf{a}$ to $\mathbf{y}$ in T must pass through the plane at infinity Π_∞, since T_1 and T_2 are different connected components of $T \setminus \Pi_\infty$. However, in this case, this line segment must pass through $\mathbf{x}$, which contradicts the assumption that S and T are disjoint.

Therefore, the sets T_2 and S' are disjoint and convex in $\mathbb{R}^3$. Theorem 11.3 of [69] ensures that there exists a plane Π separating S' from T_2. This plane is therefore disjoint from both S and T, except possibly on the plane Π_∞. However, since both S and T are assumed open, it is not possible for the plane Π to intersect S or T only on the plane at infinity.

This completes the construction of the plane disjoint from S and T. □

The previous proposition allows us to show that two open weakly convex sets may be separated by two planes.

Proposition 5 *If S and T are two disjoint open weakly convex sets in* $\mathrm{SO}(3)$*, then there exist two planes Π_1 and Π_2 such that S and T lie in different components of* $\mathrm{SO}(3) \setminus (\Pi_1 \cup \Pi_2)$.

Proof. There is a plane Π_1 that meets neither of S and T. Map this plane to infinity. Then S and T are mapped to two open convex sets in $\mathbb{R}^3$, which are therefore separable by a plane Π_2. These are the two required planes. □

Another separation property of convex sets in R^n that carries over, slightly modified to weakly convex sets in $\mathrm{SO}(3)$ is the existence of supporting planes.

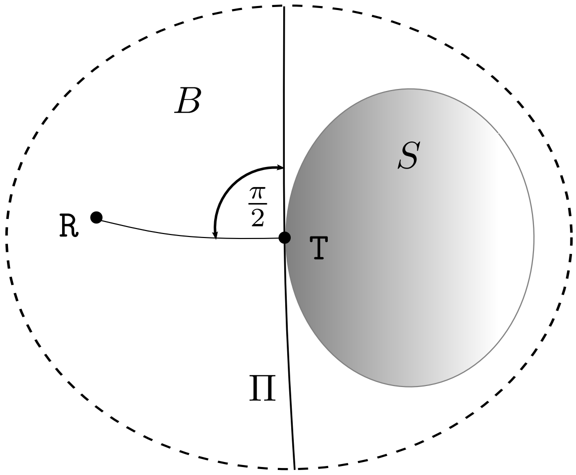


Fig. 4 *The supporting plane constructed in proposition 6.*

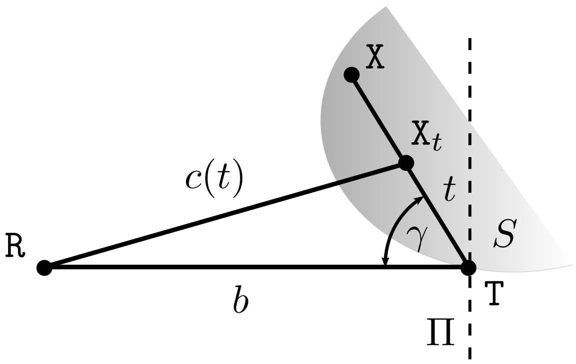


Fig. 5 *The gnomonic model used in the proof of proposition 6.*

Proposition 6 *Let S be a closed convex set in* $\mathrm{SO}(3)$*,* R *a point not in S and* T *a closest point in S to* R*. Further, let Π be the plane through* T *perpendicular to the line* RT*. Then, the plane Π divides the open ball $B = \mathring{B}(\mathtt{T}, \pi)$ into two half-balls, and S lies entirely in the closed half ball not containing* R*. Consequently, the interior of S lies in the open half ball not containing* R*.*

This situation is illustrated in fig 4. The proposition holds in a more general context than in $\mathrm{SO}(3)$, but we give a proof only for $\mathrm{SO}(3)$, using the cosine rule.

Proof. If the distance RT is equal to π, then the whole of the set S lies in the plane $\Pi(R, \pi)$, and the result is trivially true. Therefore, assume that the distance RT is less than π. Since S is convex, any point in S lies at distance less than π from T.

Via a gnomonic mapping centred at T, the ball B maps to the whole of $\mathbb{R}^3$, the set S maps to a closed bounded convex set and angles at T are preserved. We may therefore use this gnomonic model to access familiar concepts concerning sets in $\mathbb{R}^3$.

Suppose that there is a point X in S on the same side of Π as R. Since S is convex, the whole of the line TX lies in S. Furthermore, it forms an angle $\gamma < \pi/2$ with the line TR. Let $\mathtt{X}_t$ be a point on the line TX at distance t from T in the direction towards X.

Applying the cosine rule (proposition 2) to the triangle $\mathtt{RX}_t\mathtt{T}$ as shown in fig 5, we see that

$$\cos(\frac{c(t)}{2}) = \left|\cos(\frac{t}{2})\cos(\frac{b}{2}) + \sin(\frac{t}{2})\sin(\frac{b}{2})\cos(\gamma)\right|,$$

where we write $c(t)$ in recognition that the length c depends on the value of t. Since $0 \le t < \pi$ and $0 \le b < \pi$, we see that for $\gamma < \pi/2$ the expression inside the absolute value $|\cdot|$ is positive, so

$$c(t) = 2\arccos\left(\cos(\frac{t}{2})\cos(\frac{b}{2}) + \sin(\frac{t}{2})\sin(\frac{b}{2})\cos(\gamma)\right).$$

Taking derivatives with respect to t at $t = 0$, we find $dc/dt|_{t=0} = -\cos(\gamma)$, which is negative when $\gamma < \pi/2$. Thus, for sufficiently small t we have $c(t) < c(0) = b$. Thus, the point $\mathtt{X}_t$ is closer to R than the distance RT, which contradicts the assumption that T is the closest point in S to R. The conclusion is that the open half ball containing R contains no point of S, as required. □

Intersections of weakly convex sets

We now consider various properties of intersections of convex and weakly convex sets in $\mathrm{SO}(3)$ in a series of propositions. In the following discussion, we will use the language of projective geometry, speaking of lines and planes, instead of geodesics and geodesic planes. These relate to the geometric properties of $\mathrm{SO}(3)$, considered as the projective plane $\mathbb{P}^3$, in which geodesics play the role of lines in projective geometry. Note that the concept of weakly convex set is purely a property of the projective geometry of $\mathrm{SO}(3)$, viewed as a projective plane $\mathbb{P}^3$; a set S is weakly convex if any two points in S are joined by a single line segment contained in S. According to theorem 10, for any weakly convex set S there exists a plane that does not intersect S.

We consider families of convex sets B_i, indexed by i in some index set I, finite or infinite.

Proposition 7 *The intersection of a family of convex sets in* $\mathrm{SO}(3)$ *is convex or empty.*

Proof. If points x and y are in the intersection of a family of convex sets B_i then the shortest geodesic from x to y lies in each B_i, and hence in their intersection. Thus the intersection is convex. □

Proposition 8 *Consider a family of weakly convex sets B_i in* $\mathrm{SO}(3)$*. If there exists a plane Π disjoint from all of them, then their intersection is weakly convex or empty.*

Proof. Consider two points x and y in $\bigcap_{i\in I} B_i$. There exist two geodesic line segments joining x to y which together make up a complete closed geodesic. One of these line segments meets the plane Π, and hence does not lie completely inside any of the B_i. Since each B_i is weakly convex, the other line segment joining x to y must lie in B_i. Since this is true for all i, this line segment lies in the intersection of all the sets B_i, which is therefore weakly convex. □

Proposition 9 *If B is a weakly convex set in* $\mathrm{SO}(3)$ *and Π is a plane then $B \cap \Pi$ is either empty or weakly convex. Further, $B \setminus \Pi$ consists of at most two weakly convex components.*

Proof. That $B \cap \Pi$ is weakly convex unless it is empty is easily shown; we therefore turn to consider $B \setminus \Pi$.

If Π does not intersect B then $B \setminus \Pi = B$. Otherwise, according to theorem 10 there exists a plane Π' that does not intersect B, and this must be different from Π, since Π intersects B. By a suitable homography, we may map Π' to the plane at infinity. The set B maps to a convex set in $\mathbb{R}^3$ and Π to a plane in $\mathbb{R}^3$. From properties of convex sets in $\mathbb{R}^3$, the plane Π divides B into at most two parts, each of which is convex in $\mathbb{R}^3$, and hence weakly convex as a subset of $\mathrm{SO}(3)$. Note that this also covers the case where $B \setminus \Pi$ is empty. □

Proposition 10 *If B_i, $i \in I$ is a family of weakly convex sets in* $\mathrm{SO}(3)$*, then any connected component of $\bigcap_{i\in I} B_i$ is weakly convex.*

Proof. We select one B_i and choose a plane Π that it does not intersect. Then

$$\bigcap_{j\in I} B_j = \bigcap_{j\in I} (B_j \setminus \Pi).$$

Now, let $\mathbf{x}$ be a point in $\bigcap_{j\in I} B_j$, and for any $j \in I$ let B_j' be the component of $B_j \setminus \Pi$ which contains $\mathbf{x}$. It is weakly convex by proposition 9. Then $\bigcap_{j\in I} B_j'$ is the component of $\bigcap_{j\in I} B_j$ containing $\mathbf{x}$. It is weakly convex by proposition 8. Since $\mathbf{x}$ was arbitrary, every component is weakly convex. □

Proposition 11 *If B_i, $i = 1, \ldots, n$ are a finite family of weakly convex sets in* $\mathrm{SO}(3)$*, then their intersection consists of at most $\binom{n}{3} + n$ disjoint weakly convex components.*

Proof. The connected components are weakly convex by proposition 10. We simply need to estimate how many such components there are. For each B_i, select a plane Π_i that it does not intersect. The union of planes Π_i is disjoint from the intersection of the sets B_i.

Now, map the first plane Π_1 to the plane at infinity via a homography. The other $n-1$ planes divide $\mathbb{R}^3$ into convex regions V_j. Generically (if no 4 planes meet in a point and no 3 planes meet in the same line) there are $\binom{n}{3} + n$ such regions V_j, but fewer in the non-generic case [74].

Each V_j is convex in $\mathbb{R}^3$ and hence weakly convex as a subset of $\mathrm{SO}(3)$. Now,

$$V_j \cap \bigcap_{i=1}^{n} B_i = \bigcap_{i=1}^{n} (B_i \cap V_j).$$

However, each $B_i \cap V_j$ is weakly convex by proposition 8, since both B_i and V_j avoid Π_i. Similarly, the total intersection is weakly convex, since each $B_i \cap V_j$ avoids any and all of the planes Π_i.

Thus, there is at most one weakly convex component of $\bigcap_{i=1}^{n} B_i$ contained in each V_j, and hence there are not more than $\binom{n}{3} + n$ components in total. □

Convex hulls and convex basins

In the light of proposition 7 we may define the *convex hull* of a set $B \subset \mathrm{SO}(3)$ to be the minimal convex set (if one exists) that contains B. If B is not empty, and as long as there exists at least one convex set containing B, then the intersection of all such convex sets containing B is itself convex, and is therefore the convex hull of B.

Since the intersection of weakly convex sets is not generally weakly convex we cannot define a weakly convex hull of a set of points in the same way. For example, a line segment of length less than 2π is weakly convex, but the intersection of two line segments of length $3/2\pi$ arranged suitably on a single line will not be connected and hence not weakly convex. This is easily pictured thinking of lines (closed geodesics) as circles. Under certain circumstances, however, there will exist a smallest weakly convex set containing a set B. We therefore make the following definition.

Definition 2 Let S be a set in $\mathrm{SO}(3)$ and H a weakly convex set containing S. If H is a subset of any other weakly convex set H' that contains S, then we say that the weakly convex hull of S exists, and is H.

Thus, H is the minimal weakly convex set containing S, if such a minimal set exists. Note that not every set has a weakly convex hull, even if it is contained in some weakly convex set. The empty set has no weakly convex hull since the empty set is not considered to be weakly convex.

We list some simple properties of weakly convex hulls.

Proposition 12 *A nonempty set S in* $\mathrm{SO}(3)$ *has a weakly convex hull if and only if the intersection of all weakly convex sets H_i containing S consists of a single connected component. This component is the weakly convex hull.*

The proof is immediate.

Sets with weakly convex hulls can be characterized simply in terms of connectivity. A nonempty set S may be called *convex-connected* if whenever S is contained in the disjoint union of two open weakly convex sets, $S \subset H_1 \cup H_2$, then either $S \cap H_1$ or $S \cap H_2$ is empty. Note that this is analogous to the usual definition of a connected set; in fact every connected set is convex-connected. It may seem more appropriate to say that S is *weakly convex-connected*, but this seems too verbose, so we choose this terminology.

Proposition 13 *A nonempty set S in $\mathrm{SO}(3)$ has a weakly convex hull if and only if it is contained in some weakly convex set and is convex-connected.*

Proof. Suppose that S is convex-connected and contained in the weakly convex set B. Let Π_S be a plane that does not intersect B (theorem 10) and hence does not intersect S. We define $H = \bigcap_i B_i$ where B_i runs over all weakly convex sets containing S. If we can show that H is itself weakly convex, then it is the weakly convex hull of S. This will be accomplished by showing that

$$H = \bigcap_i B_i = \bigcap_i B'_i \tag{29}$$

where B'_i is a weakly convex subset of B_i and $B'_i \cap \Pi_S = \emptyset$. In this case H is weakly convex according to proposition 8.

To this end, let B_i be such a weakly convex set containing S. The plane Π_S divides B_i into at most two weakly convex sets, $B_i \setminus \Pi_S = B_i^1 \cup B_i^2$ (proposition 9), where B_i^2 may be empty. Since $B_i \setminus \Pi_S$ contains S, the other component B_i^1 will then be nonempty. Now let Π_i be a plane not intersecting B_i. Then $\mathrm{SO}(3) \setminus (\Pi_S \cup \Pi_i)$ is a union of two disjoint open weakly convex sets, and it contains S. Therefore, S is contained in one of these two sets, since S is assumed to be convex-connected. Furthermore, since either B_i^2 is empty or B_i^1 and B_i^2 lie in different sets, it follows that $S \subset B_i^1$ or $S \subset B_i^2$. In particular, we may replace B_i in (29) by B'_i, where B'_i is the component of $B_i \setminus \Pi_S$ containing S. This completes the demonstration that S has a weakly convex hull.

Conversely, suppose that S has a weakly convex hull H, which is therefore a weakly convex set containing S and is the intersection of all weakly convex sets containing S. Let H_1 and H_2 be two disjoint weakly convex open sets with $S \subset H_1 \cup H_2$. Let Π_1 and Π_2 be two planes such that H_1 and H_2 are in different components of $\mathrm{SO}(3) \setminus (\Pi_1 \cup \Pi_2)$. These planes exist according to proposition 5. Then $\mathrm{SO}(3) \setminus \Pi_1$ and $\mathrm{SO}(3) \setminus \Pi_2$, are both weakly convex sets containing S. It follows that H is disjoint from both Π_1 and Π_2. Suppose neither $S \cap H_1$ nor $S \cap H_2$ is empty. Then S, and hence H contains points from both components of $\mathrm{SO}(3) \setminus (\Pi_1 \cup \Pi_2)$, so H cannot be connected. This is a contradiction since H is weakly convex, and leads to the conclusion that S is contained completely in one of the two sets H_1 or H_2. Hence S is convex-connected. □

As a simple corollary of this result, a connected set S contained in some weakly convex set B has a weakly convex hull.

Convex basins. We now turn to the study of *convex basins* of sets S in $\mathrm{SO}(3)$. These will be important in defining the domain of convexity of sums of distance functions defined on $\mathrm{SO}(3)$, in section 5.

For $\mathbf{x} \in \mathrm{SO}(3)$, define $\Pi(\mathbf{x})$ to be the plane consisting of all points at distance π from $\mathbf{x}$.

Let S be a set in $\mathrm{SO}(3)$. We define the set

$$S^\natural = \bigcap_{\mathbf{x} \in S} \mathring{B}(\mathbf{x}, \pi) = \mathrm{SO}(3) \setminus \bigcup_{\mathbf{x} \in S} \Pi(\mathbf{x}),$$

which will be called the *convex basin* of S. The following implications are easily demonstrated for a point $\mathbf{x}$ and set S in $\mathrm{SO}(3)$, following directly from the definition of $S^\natural$.

$$\mathbf{x} \in S^\natural \quad \Leftrightarrow \quad \Pi(\mathbf{x}) \cap S = \emptyset \quad \Leftrightarrow \quad S \subset \mathring{B}(\mathbf{x}, \pi), \tag{30}$$

$$\mathbf{x} \in S \quad \Rightarrow \quad \Pi(\mathbf{x}) \cap S^\natural = \emptyset \quad \Leftrightarrow \quad S^\natural \subset \mathring{B}(\mathbf{x}, \pi). \tag{31}$$

Note that the implication on the left in (31) is not bidirectional; for example, $\Pi(\mathbf{y})^\natural = \emptyset$ for any $\mathbf{y} \in \mathrm{SO}(3)$.

We give some properties of convex basins.

Proposition 14 *If S is a weakly convex set then so is $S^\natural$; in particular, $S^\natural$ is connected.*

Proof. Consider two points $\mathbf{y}_0$ and $\mathbf{y}_1$ in $S^\natural$, lying on a line L and dividing L into two line segments L_0 and L_1. We show that one of the line segments L_i lies entirely in $S^\natural$. Assume the contrary; thus for $i = 1, 2$, there exist points $\mathbf{x}_0 \in L_0$ and $\mathbf{x}_1 \in L_1$ with $\mathbf{x}_i \notin S^\natural$.

Therefore, by (30) there exist points $\mathbf{x}'_i \in S$ such that $\mathbf{x}'_i \in \Pi(\mathbf{x}_i)$ or, equivalently, such that $\mathbf{x}_i \in \Pi(\mathbf{x}'_i)$. Since S is weakly convex, there exist points $\mathbf{x}'_t \in S$, for $t \in [0, 1]$ tracing out the line segment from $\mathbf{x}'_0$ to $\mathbf{x}'_1$. For each t, let $\mathbf{x}_t = L \cap \Pi(\mathbf{x}'_t)$. Note that this intersection must be a single point, since $\Pi(\mathbf{x}'_t)$ does not contain the line L because $\mathbf{y}_i \in S^\natural$ lies on L. Also, for $t = 0$ and $t = 1$ we recover our previous points $\mathbf{x}_0$ and $\mathbf{x}_1$, respectively. Then $\mathbf{x}'_t \in \Pi(\mathbf{x}_t) \cap S$ and $\mathbf{x}_t \notin S^\natural$ by (30). Furthermore, $\mathbf{x}_t$ traces out a path from $\mathbf{x}_0$ to $\mathbf{x}_1$ on L. This path must pass through $\mathbf{y}_0$ or $\mathbf{y}_1$, contradicting the assumption that $\mathbf{y}_0, \mathbf{y}_1 \in S^\natural$.

On the other hand, the whole line $L = L_1 \cup L_2$ cannot lie in $S^\natural$, since if $\mathbf{x}$ is any point in S, then $\Pi(\mathbf{x}) \cap L$ is nonempty (a plane and a line must meet). Thus some point in L is not in $S^\natural$, unless S is empty. □

Proposition 15 *If S has a weakly convex hull H, then $S^\natural = H^\natural$; in particular, $S^\natural$ is weakly convex.*

Proof. Since $S \subset H$, it follows easily that $H^\natural \subset S^\natural$. Now, let $\mathbf{x} \in S^\natural$, so $S \subset \mathring{B}(\mathbf{x}, \pi)$ by (30). This is a weakly convex set containing S. Since H is the minimal weakly convex set containing S, it follows that $H \subset \mathring{B}(\mathbf{x}, \pi)$, and so $\mathbf{x} \in H^\natural$ (again by (30)). Hence, $S^\natural \subset H^\natural$, and the result follows. □

Proposition 16 *If S is connected, then so is $S^\natural$.*

Proof. Since S is connected, it is convex-connected. If there exists some plane Π disjoint from S, then proposition 13 shows that S has a weakly convex hull, so by proposition 15, $S^\natural$ is weakly convex, hence connected.

On the other hand if each plane Π meets S, consider a point $\mathbf{x} \in \mathrm{SO}(3)$. Since $\Pi(\mathbf{x}) \cap S \neq \emptyset$, it follows (from (30)) that $\mathbf{x} \notin S^\natural$. Thus $S^\natural$ is empty, and hence connected. □

Proposition 17 *If S is an open set then $S^\natural$ is closed. If S is closed, then $S^\natural$ is open.*

Proof. It is easily seen that if B is an open ball then $B^\natural$ is a closed ball. Now if S is open, then it is the union of open balls B_i. Consequently, $S^\natural = \bigcap_i B_i^\natural$, which is closed.

Next, suppose S is closed and consider a convergent sequence of points $\mathbf{x}_i$ in $\mathrm{SO}(3) \setminus S^\natural = \bigcup_{\mathbf{y}\in S} \Pi(\mathbf{y})$. We wish to show that their limit point $\mathbf{x}_{\mathrm{lim}}$ is also in $\mathrm{SO}(3) \setminus S^\natural$. This would imply that $\mathrm{SO}(3) \setminus S^\natural$ is closed, so $S^\natural$ is open.

We choose points $\mathbf{y}_i$ in S such that $\mathbf{x}_i \in \Pi(\mathbf{y}_i)$. Since S is closed, hence compact, there exists a convergent subsequence of $\mathbf{y}_i$ converging to a point $\mathbf{y}_{\mathrm{lim}}$ in S. Select a value $\varepsilon > 0$. There exist points $\mathbf{y}_i$ and $\mathbf{x}_i$ such that $d(\mathbf{y}_i, \mathbf{y}_{\mathrm{lim}}) < \varepsilon$, $d(\mathbf{x}_i, \mathbf{x}_{\mathrm{lim}}) < \varepsilon$, and by definition $d(\mathbf{y}_i, \mathbf{x}_i) = \pi$. By the triangle inequality, $\pi - 2\varepsilon < d(\mathbf{x}_{\mathrm{lim}}, \mathbf{y}_{\mathrm{lim}}) < \pi + 2\varepsilon$. Since ε is arbitrary, it follows that $d(\mathbf{x}_{\mathrm{lim}}, \mathbf{y}_{\mathrm{lim}}) = \pi$. Since $\mathbf{y}_{\mathrm{lim}} \in S$, it follows that $\mathbf{x}_{\mathrm{lim}} \in \mathrm{SO}(3) \setminus S^\natural$. □

The following result shows that the relationship $S \leftrightarrow S^\natural$ is a dual relationship between open and closed weakly convex sets.

Proposition 18 *If S is an open or closed weakly convex set then $S^{\natural\natural} = S$.*

Proof. If $\mathbf{x} \in S$ then $\Pi(\mathbf{x}) \cap S^\natural = \emptyset$, by (31). Then by (30), $\mathbf{x} \in S^{\natural\natural}$, so S is contained in $S^{\natural\natural}$. To show the inverse inclusion, let $\mathbf{x}$ be a point not in S. As remarked in proposition 3, there exists a plane through $\mathbf{x}$ that does not intersect S. Let this plane be $\Pi(\mathbf{x}')$. Then $\mathbf{x}' \in S^\natural$ (by (30)), and so $\Pi(\mathbf{x}') \cap S^{\natural\natural} = \emptyset$ (by (31)). In particular $\mathbf{x} \notin S^{\natural\natural}$. □

Proposition 19 *If S is contained in a convex set H, then H is contained in a single connected component of $S^\natural$. In particular, if S is itself convex, then $S^\natural$ is a weakly-convex set containing S.*

Proof. Since the distance between two points in H is less than π, no plane $\Pi(\mathbf{x}), \mathbf{x} \in S$ will intersect with H. Consequently, $\bigcup_{\mathbf{x}\in S} \Pi(\mathbf{x})$ is disjoint from H, and H lies fully inside $S^\natural = \mathrm{SO}(3) \setminus \bigcup_{\mathbf{x}\in S} \Pi(\mathbf{x})$. Since H is connected it lies within a single connected component of this set. □

Examples. Let S be the closed ball $B(\mathtt{S}, r)$, with $r < \pi$. Then $S^\natural$ is the open ball $\mathring{B}(\mathtt{S}, \pi - r)$. Similarly, if S is the open ball $\mathring{B}(\mathtt{S}, r)$ with $r \leq \pi$, then $S^\natural$ is the closed ball $B(\mathtt{S}, \pi - r)$.

In particular when $r = \pi/2$ and $S = \mathring{B}(\mathtt{S}, \pi/2)$, then $S^\natural = B(\mathtt{S}, \pi/2)$. This is a special case of proposition 19.

Convex functions in SO(3)

Convex functions can be defined as in $\mathbb{R}^n$, except that geodesic curves in SO(3) take the place of straight lines joining two points in $\mathbb{R}^n$. To make this explicit, we need the following terminology, requiring geodesic curves to be parametrized to have constant speed.

A *geodesic curve* in SO(3) is a *constant speed* path along a geodesic. Here, we think of speed as being defined in terms of the angle metric in SO(3), but either of the other metrics d_{chord} or d_{quat} can be used instead, since they result in the same path length (except for scale).

Definition 3 Consider a function $f : U \to \mathbb{R}$ defined on a weakly convex subset U of SO(3). Let $\mathbf{x}_0, \mathbf{x}_1 \in U$ and let $g : [0, 1] \to U$ be a geodesic curve from $\mathbf{x}_0$ to $\mathbf{x}_1$ in U, such that $g(0) = \mathbf{x}_0$ and $g(1) = \mathbf{x}_1$. The function f is called *convex*, if for any such $\mathbf{x}_0$, $\mathbf{x}_1$ and g, we have an inequality

$$f(g(\lambda)) \leq (1 - \lambda) f(\mathbf{x}_0) + \lambda f(\mathbf{x}_1)$$

for all $\lambda \in [0, 1]$. The function is called *strictly convex* if this inequality is strict for all $\lambda \in (0, 1)$ whenever $\mathbf{x}_0 \neq \mathbf{x}_1$.

Various properties of convex functions hold true, just as with convex functions in $\mathbb{R}^n$.

Proposition 20 *1. The sum of convex (or strictly convex) functions defined on a weakly convex region U is convex (respectively, strictly convex) .*
2. A strictly convex function defined on a weakly convex set has at most a single local minimum, which is therefore the global minimum; for convex functions (even if they are not strictly convex), any local minimum is a global minimum and the minima form a weakly convex set on which the function is constant.

The proof is the same as for convex functions in $\mathbb{R}^n$.

Convexity of functions can be defined locally through computing the second derivative of their restriction along geodesic paths through a point.

Definition 4 A function $f : \mathrm{SO}(3) \to \mathbb{R}$ is *locally convex* at a point $\mathtt{R}_0 \in \mathrm{SO}(3)$ if for any constant speed geodesic path $\gamma : [-1, 1] \to \mathrm{SO}(3)$, with $\gamma(0) = \mathtt{R}_0$ the function $f \circ \gamma(t) = f(\gamma(t))$ has non-negative second derivative at $t = 0$. It is *locally strictly convex* at $\mathtt{R}_0$ if any such $f \circ \gamma(t)$ has positive second derivative at $t = 0$.

The connection between local convexity and convexity is as follows.

Proposition 21 *If $f : \mathrm{SO}(3) \to \mathbb{R}$ is smooth and locally convex (or strictly convex) at each point in a weakly convex set U, except possibly at isolated global minima of f, then it is convex (respectively, strictly convex) in U. If $f : \mathrm{SO}(3) \to \mathbb{R}$ is smooth but not locally convex at some point then it is not convex in any non-trivial ball around that point.*

Next we investigate when the function $d(\mathtt{S}, \mathtt{R})$ defined for two rotations is a convex function of S (for fixed R).

Theorem 11 (Convexity of metrics) *Consider the function $f(\mathtt{S}) = d(\mathtt{S}, \mathtt{R})^p$ for a fixed rotation R, a metric $d(\cdot, \cdot)$, and an exponent p. The function is convex, or strictly convex, as a function of S in the following cases.*

1. *$d_\angle(\cdot, \mathtt{R})$ is convex on the set $\mathring{B}(\mathtt{R}, \pi)$.*
2. *$d_{\mathrm{chord}}(\cdot, \mathtt{R})$ is not convex on any non-trivial ball around R.*
3. *$d_{\mathrm{quat}}(\cdot, \mathtt{R})$ is not convex on any non-trivial ball around R.*
4. *$d_\angle(\cdot, \mathtt{R})^2$ is strictly convex on the set $\mathring{B}(\mathtt{R}, \pi)$.*
5. *$d_{\mathrm{chord}}(\cdot, \mathtt{R})^2$ is strictly convex on the set $B(\mathtt{R}, \pi/2)$.*
6. *$d_{\mathrm{quat}}(\cdot, \mathtt{R})^2$ is strictly convex on the set $\mathring{B}(\mathtt{R}, \pi)$.*

Compare these results to the graphs in fig 2 in section 4. From these graphs, parts 2 and 3 of the theorem are evident. It is also clear that $d_\angle(\cdot, \mathtt{R})$ is not strictly convex anywhere. The other parts of the theorem are obtained by direct computation of second derivatives. Details of how these values are computed and a table of Hessians and gradients are found in table 3 in the following appendix.

Two geometric lemmas

The following two lemmas are used in the proof of Theorem theorem 5.

Lemma 8 (Pumping lemma.) *Let B be a closed convex subset of $\mathrm{SO}(3)$ then there exists a larger closed convex subset $\hat{B}$ of $\mathrm{SO}(3)$ such that all points of B lie in the interior of $\hat{B}$. Furthermore, the intersection of all such sets $\hat{B}$ is equal to B.*

Proof If B is a closed convex set, then its diameter must be strictly less than π. Let ε be a number such that $\mathrm{diameter}(B) + 4\varepsilon < \pi$. Now, let Γ be the gnomonic map based at some point in B. This takes B to a closed bounded convex set $\Gamma(B)$ in $\mathbb{R}^3$. Let $N_\varepsilon(\Gamma(B))$ be an ε-neighbourhood of $\Gamma(B)$, that is, the union of closed balls of radius ε centred on points of $\Gamma(B)$. This is a closed convex set in $\mathbb{R}^3$ containing $\Gamma(B)$ in its interior. Let $B' = \Gamma^{-1}(N_\varepsilon(\Gamma(B)))$, which is a closed weakly convex set in $\mathrm{SO}(3)$. To show that B' is convex, it remains to show that the diameter of B' is less than π.

The gnomonic map expands distances. More exactly, elementary trigonometry shows that $\|\Gamma(\mathtt{R}) - \Gamma(\mathtt{S})\| > \alpha = d_\angle(\mathtt{R}, \mathtt{S})/2$, where α is the angle between R and S on the unit quaternion sphere. In particular, the inverse image under Γ^{-1} of a closed ball of radius ε in $\mathbb{R}^3$ is a set of radius less than 2ε in $\mathrm{SO}(3)$. It follows using the triangle inequality that the diameter of B' is no more than $\mathrm{diameter}(B) + 4\varepsilon < \pi$.

Lemma 9 *Theorem 5 is true in the special case where B is a closed convex set and the rotations $\mathtt{R}_i$ lie in the interior of B.*

Proof Let B be a closed convex set containing all $\mathtt{R}_i$ in its interior and let R be a point not in B. We will show that R cannot be the point that minimizes the cost $C_f(\mathtt{R})$ by explicitly computing a point $\mathtt{R}'$ with lesser cost. Since B is compact, there exists a point $\mathtt{T} \in B$ that minimizes the distance to R. There may be more than one such point T, but we take any one. We observe first that $d_\angle(\mathtt{R}, \mathtt{T}) < \pi$, since if this is not true, then T and hence every point in B must be at distance π (the maximum possible distance) from R. In this case B lies in the plane at distance π from R, and hence has empty interior, contrary to assumption.

Now, if we were in $\mathbb{R}^n$, we could argue that $d_\angle(\mathtt{T}, \mathtt{R}_i) < d_\angle(\mathtt{R}, \mathtt{R}_i)$, for any point $\mathtt{R}_i \in B$, but this is not true in $\mathrm{SO}(3)$. Instead we find a point $\mathtt{R}'$ such that $d_\angle(\mathtt{R}', \mathtt{R}_i) < d_\angle(\mathtt{R}, \mathtt{R}_i)$, and hence $d_i(\mathtt{R}') < d_i(\mathtt{R})$, which proves that R is not the point that minimizes C_f.

The point $\mathtt{R}'$ is constructed as follows. Consider the minimal geodesic from R to T and continue it beyond T by the same distance to a point $\mathtt{R}'$. Thus $d_\angle(\mathtt{T}, \mathtt{R}) = d_\angle(\mathtt{T}, \mathtt{R}') < \pi$. We do not claim that $\mathtt{R}' \in B$, or that $\mathtt{R}'$ minimizes the cost function. Next, consider the plane Π passing through T perpendicular to the geodesic from R to T. The configuration described here satisfies the hypotheses of proposition 6.

Now, we consider the gnomonic projection Γ centred at T. Since the diameter of B is less than π, and $\mathtt{T} \in B$, the whole of B is mapped to a bounded convex set in $\mathbb{R}^3$. Similarly, the shortest geodesic from R to T maps to a bounded line segment in $\mathbb{R}^3$, not meeting the interior of $\Gamma(B)$, and the plane Π maps to a plane in $\mathbb{R}^3$. Since the gnomonic map preserves angles at the base point, $\Gamma(\Pi)$ is perpendicular to the line from $\Gamma(\mathtt{R})$ to $\Gamma(\mathtt{T})$.

According to proposition 6, the plane $\Gamma(\Pi)$ separates $\mathbb{R}^3$ into two half-spaces, with the interior of $\Gamma(B)$ and $\Gamma(\mathtt{R}')$ lying in one half space, and $\Gamma(\mathtt{R})$ in the other. This is shown in fig 6. For a point $\mathtt{S} \in \mathring{B}$, we claim that the angle RTS is greater than $\pi/2$. This is obvious for the corresponding points in $\mathbb{R}^3$ since $\Gamma(\mathtt{S})$ is separated from $\Gamma(\mathtt{R})$ by the plane $\Gamma(\Pi)$ which passes through $\Gamma(\mathtt{T})$. Since the gnomonic projection preserves angles at the base point, the claim is valid

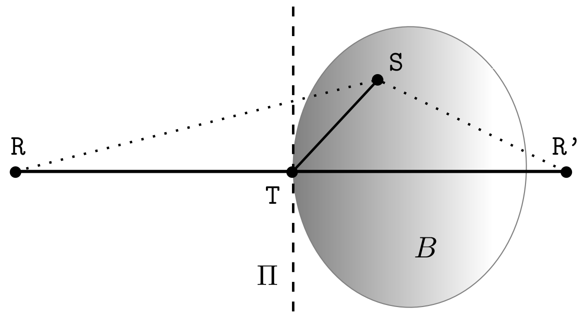


Fig. 6 *The supporting plane in the gnomonic picture.*

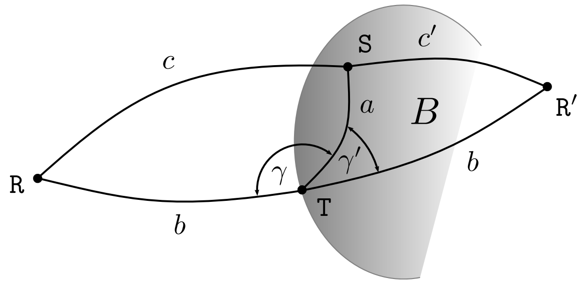


Fig. 7 *Notation used in proving that* $c = d_\angle(\mathtt{S},\mathtt{R}) > c' = d_\angle(\mathtt{S},\mathtt{R}')$.

in $\mathrm{SO}(3)$. Since $\mathtt{R}$, $\mathtt{T}$ and $\mathtt{R}'$ lie on a single geodesic, it follows that the angle $\mathtt{R}'\mathtt{TS} < \pi/2$.

To complete the proof, it is sufficient to show that $d_\angle(\mathtt{S},\mathtt{R}') < d_\angle(\mathtt{S},\mathtt{R})$. Note that we can not appeal to the gnomonic projection to demonstrate this claim, which would be obvious in $\mathbb{R}^3$, since the gnomonic projection does not preserve lengths. Furthermore, we do not know whether the shortest geodesics from $\mathtt{S}$ to $\mathtt{R}$ or $\mathtt{R}'$ cross the plane at infinity in the gnomonic projection or not.

To prove the claim, we appeal to the cosine rule (13) to compute geodesic lengths in $\mathrm{SO}(3)$. We use notation as shown in fig 7, where $c = d_\angle(\mathtt{S},\mathtt{R}) \leq \pi$ and $c' = d_\angle(\mathtt{S},\mathtt{R}') \leq \pi$. Since $\gamma + \gamma' = \pi$, it follows that $\cos(\gamma) = -\cos(\gamma')$. Then applying the cosine rule, we find

$$\cos(\frac{c}{2}) = \left|\cos(\frac{a}{2})\cos(\frac{b}{2}) - \sin(\frac{a}{2})\sin(\frac{b}{2})\cos(\gamma')\right|$$
$$\cos(\frac{c'}{2}) = \left|\cos(\frac{a}{2})\cos(\frac{b}{2}) + \sin(\frac{a}{2})\sin(\frac{b}{2})\cos(\gamma')\right|$$

Now, $0 < a < \pi$ and $0 < b < \pi$, so $\sin(a/2)\sin(b/2) > 0$, and $\cos(a/2)\cos(b/2) > 0$. Furthermor $\cos(\gamma') > 0$, since $\gamma' < \pi/2$. It follows easily that $\cos(c'/2) > \cos(c/2) \geq 0$, so $c' < c$ as required.

Appendix – gradients and Hessians

Given a function $f : \mathrm{SO}(3) \to \mathbb{R}$, we wish to define and calculate the gradient and Hessian of this function. These entities may be expressed in terms of the exponential map at the point of interest. Let $\exp_{\mathtt{R}} : \mathbb{R}^3 \to \mathrm{SO}(3)$ be the exponential map at a point $\mathtt{R} \in \mathrm{SO}(3)$, defined by $\exp_{\mathtt{R}}[\mathbf{v}]_\times = \mathtt{R}\exp[\mathbf{v}]_\times$. The gradient and Hessian of the function f at the point $\mathtt{R}$ are defined as the gradient and Hessian (the matrix of second derivatives) of the function $f \circ \exp_{\mathtt{R}} : \mathbb{R}^3 \to \mathbb{R}$, evaluated at $\mathbf{v} = \mathbf{0}$.

This definition corresponds with the notion of Riemannian gradient and Hessian, which are defined on the tangent space $T_{\mathtt{R}}(\mathrm{SO}(3))$ to $\mathrm{SO}(3)$ at the point $\mathtt{R}$. In this more abstract context, the Hessian is a quadratic form defined on the tangent space. If we identify $\mathbb{R}^3$ with its standard Euclidean basis as the tangent space, this quadratic form is represented by the symmetric second derivative matrix defined here.

We have defined the concept of convexity of a function defined on $\mathrm{SO}(3)$ in terms of the values of the function along geodesics.

Theorem 12 *If the Hessian of a function* $f : \mathrm{SO}(3) \to \mathbb{R}$ *is positive semi-definite at a point* $\mathtt{R}_0 \in \mathrm{SO}(3)$*, then* f *is locally convex at* $\mathtt{R}_0$*. If the Hessian is positive definite, then the function is locally strictly convex.*

Proof. Let $\gamma : \mathbb{R} \to \mathrm{SO}(3)$ be a constant speed geodesic path with $\gamma(0) = \mathtt{R}_0$. We may pull γ back to a path $\tilde{\gamma} : \mathbb{R} \to \mathbb{R}^3$ such that $\gamma = \exp_{\mathtt{R}_0} \circ \tilde{\gamma}$. To show that f is locally convex, we need to show that $f \circ \gamma(t) = f \circ \exp_{\mathtt{R}_0} \circ \tilde{\gamma}(t)$ has non-negative second derivative at $t = 0$. However, the second derivative may be written as $\mathbf{v}^\top \mathtt{H} \mathbf{v}$, where $\mathtt{H}$ is the Hessian of $f \circ \exp_{\mathtt{R}_0}$ and $\mathbf{v} = \tilde{\gamma}'(0)$ is the derivative of $\tilde{\gamma}$. If the Hessian is positive definite (or semi-definite), this is positive (non-negative) as required. □

Thus, to show that a function on $\mathrm{SO}(3)$ is convex, it is sufficient to show that its Hessian is positive definite, except possibly at isolated local minima.

Gradient and Hessian of distance functions. Consider $\mathtt{S} \in \mathrm{SO}(3)$ and let $f(\mathtt{R}) = d^p(\mathtt{R},\mathtt{S})$ where $d(\cdot,\cdot)$ is some bi-invariant metric defined on $\mathrm{SO}(3)$. By definition, $\mathtt{H}_f$ is the Hessian of the function

$$\tilde{f}(\mathbf{x}) = f(\mathtt{R}\exp[\mathbf{x}]_\times) = d^p(\exp[\mathbf{x}]_\times, \mathtt{R}^\top\mathtt{S}) \;.$$

Define $\theta = d_\angle(\exp[\mathbf{x}]_\times, \mathtt{R}^\top\mathtt{S})$ and let $\mathtt{R}^\top\mathtt{S}$ be a rotation through angle θ_0 about unit axis $\hat{\mathbf{w}}$. Then, using the cosine rule (13), we may write

$$\cos(\frac{\theta}{2}) = \cos(\frac{\|\mathbf{x}\|}{2})\cos(\frac{\theta_0}{2}) + \sin(\frac{\|\mathbf{x}\|}{2})\sin(\frac{\theta_0}{2})\frac{\langle \mathbf{x},\,\hat{\mathbf{w}}\rangle}{\|\mathbf{x}\|_2} \;.$$

Since we wish to take derivatives up to second order, we may replace this by its second-order approximation, yielding

$$\theta \approx 2\arccos\left(\left(1 - \frac{\|\mathbf{x}\|^2}{8}\right)\cos(\frac{\theta_0}{2}) + \frac{1}{2}\sin(\frac{\theta_0}{2})\,\langle \mathbf{x},\,\hat{\mathbf{w}}\rangle\right).$$

Now, we define $f(\mathtt{R}) = d^p(\mathtt{R},\mathtt{S}) = g(d_\angle(\mathtt{R},\mathtt{S})) = g(\theta)$, for some function g. The various metrics being considered can all be expressed in this way for suitable functions g (see table 2). Taking first derivatives using the chain rule gives

$$\frac{\partial \tilde{f}}{\partial x_i} = \frac{\partial g}{\partial \theta}\frac{\partial \theta}{\partial x_i} \quad \text{or} \quad \nabla_f = g'(\theta_0)\,\nabla_\theta \ .$$

Evaluating at the point $\mathbf{x} = 0$ gives the gradient

$$\nabla_f = -g'(\theta_0)\hat{\mathbf{w}} \ .$$

In interpreting this, note that $\mathtt{R}\exp[t\hat{\mathbf{w}}]_\times = \exp_\mathtt{R}[t\hat{\mathbf{w}}]_\times$ is a geodesic from $\mathtt{R}$ when $t = 0$ to $\mathtt{S}$ when $t = 1$. Thus, as a vector in the tangent space at $\mathtt{R}$, the unit vector $\hat{\mathbf{w}}$ may be viewed as the direction from $\mathtt{R}$ to $\mathtt{S}$. The gradient points directly away from $\mathtt{S}$, in the direction of greatest increasing distance.

Similarly, taking second derivatives using the chain and product rules leads to

$$\frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j} = \frac{\partial^2 g}{\partial \theta^2}\frac{\partial \theta}{\partial x_i}\frac{\partial \theta}{\partial x_j} + \frac{\partial g}{\partial \theta}\frac{\partial^2 \theta}{\partial x_i \partial x_j}$$

or

$$\mathtt{H}_f = g''(\theta_0)\,\nabla_\theta\,\nabla_\theta{}^\top + g'(\theta_0)\mathtt{H}_\theta \ .$$

From this it is straight-forward to compute the Hessian. The result is

$$\mathtt{H}_f = g''(\theta_0)\,\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i{}^\top + g'(\theta_0)\frac{\cot(\theta_0/2)}{2}\,(\mathtt{I} - \hat{\mathbf{w}}_i\hat{\mathbf{w}}_i{}^\top) \ .$$

Note that both $\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i{}^\top$ and $\mathtt{I} - \hat{\mathbf{w}}_i\hat{\mathbf{w}}_i{}^\top$ can be diagonalized simultaneously to $\text{diag}(1,0,0)$ and $\text{diag}(0,1,1)$. Thus, the Hessian may be transformed orthogonally (but differently for each i) to the form

$$\mathtt{H}_f \approx g''(\theta_0)\text{diag}(1,0,0) + g'(\theta_0)\frac{\cot(\theta_0/2)}{2}\,\text{diag}(0,1,1) \ .$$

In particular, the Hessian is positive definite exactly when both the derivatives of g are positive. We can apply this formula with different functions g to obtain the results in table 3.

Conjugate distance function. Given rotations $\mathtt{R}_i$ and $\mathtt{L}_i$, we consider the function $\mathtt{S} \mapsto d^p(\mathtt{R}_i\mathtt{S}, \mathtt{S}\mathtt{L}_i)$. We wish to compute the gradient and Hessian of this function. For simplicity, we will compute these quantities at the point $\mathtt{S} = \mathtt{I}$, and see later that the general case is easily derived from this special case. Setting $\mathtt{S} = \exp[\mathbf{x}]_\times$, the gradient and Hessian are defined as the gradient and Hessian of $d^p(\mathtt{R}_i\exp[\mathbf{x}]_\times, \exp[\mathbf{x}]_\times\mathtt{L}_i)$ with respect to the vector $\mathbf{x}$.

Let $\mathbf{r}_i$, $\mathbf{l}_i$ and $\mathbf{s}$ be corresponding quaternion representations, chosen to lie in the upper quaternion hemisphere. Let $\theta_i = d_\angle(\mathtt{R}_i\mathtt{S}, \mathtt{S}\mathtt{L}_i)$ and define $C = \cos(\theta_i/2)$. Then, C may be written in terms of the quaternion inner product

$$C = \langle \mathbf{r}_i \cdot \mathbf{s},\, \mathbf{s} \cdot \mathbf{l}_i \rangle \ .$$

Let the quaternion representations of $\mathtt{R}_i$ and $\mathtt{L}_i$ be $\mathbf{r}_i = (r_0, \mathbf{r}'_i)$ and $\mathbf{l}_i = (l_0, \mathbf{l}'_i)$. The quaternion representation of $\mathtt{S} = \exp[\mathbf{x}]_\times$ is $(\cos(\|\mathbf{x}\|/2), \sin(\|\mathbf{x}\|/2)\mathbf{x}/\|\mathbf{x}\|)$, which, as above, we may replace by its second-order approximation $\mathbf{s} = (1 - \|\mathbf{x}\|^2/8, \mathbf{x}/2)$. Now, we may compute the inner product $C = \langle \mathbf{r}_i \cdot \mathbf{s},\, \mathbf{s} \cdot \mathbf{l}_i \rangle$, and differentiate with respect to $\mathbf{x}$. The results for the gradient and Hessian of C are

$$\nabla_C = \mathbf{l}'_i \times \mathbf{r}'_i \ , \tag{32}$$

and

$$\mathtt{H}_C = (\mathbf{l}'_i\mathbf{r}'^\top_i + \mathbf{r}'_i\mathbf{l}'^\top_i)/2 - \langle \mathbf{l}'_i,\, \mathbf{r}'_i \rangle\, \mathtt{I} \ . \tag{33}$$

Note that $\mathbf{r}'_i$ and $\mathbf{l}'_i$ are vectors of length $\sin(\theta^r_i/2)$ and $\sin(\theta^l_i/2)$, where θ^r_i and θ^l_i are the respective rotation angles of $\mathtt{R}_i$ and $\mathtt{L}_i$. Hence, the above formulas may easily be rewritten in terms of the unit rotation axes of the rotations, by multiplying by weights $w_i = \sin(\theta^r_i/2)$ resp. $\sin(\theta^l_i/2)$. The eigenvalues of $\mathtt{H}_C$ may be easily computed, and expressed in the form $(w_i\cos(\alpha_i/2), w_i(\cos(\alpha_i) - 1), w_i(\cos(\alpha_i) + 1))$ where α_i is the angle between the axes of $\mathtt{R}_i$ and $\mathtt{L}_i$. Hence, the Hessian has at least one negative eigenvalue, unless $\alpha_i = 0$, when it has two positive and one zero eigenvalue.

Let $d^p(\mathtt{R}_i\mathtt{S}, \mathtt{S}\mathtt{L}_i)$ be written as $g(C)$ for some appropriate function g. For example, since $C = \cos(\theta/2)$, we have $d_{\text{quat}}(\cdot,\cdot)^2 = 4\sin^2(\theta/4) = 2(1 - C)$ and $d_{\text{chord}}(\cdot,\cdot)^2 = 8\sin^2(\theta/2) = 8(1 - C^2)$. The gradient and Hessian may then be expressed as in table 4.

References

1. P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, Princeton, NJ, 2008. With a foreword by Paul Van Dooren.
2. B. Afsari. Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011.
3. M. Agrawal. A Lie algebraic approach for consistent pose registration for general euclidean motion. In *International Conference on Intelligent Robots and Systems*, pages 1891–1897, Oct 2006.
4. S. L. Altmann. *Rotations, quaternions, and double groups*. Oxford Science Publications. The Clarendon Press Oxford University Press, New York, 1986.
5. N. Asgharbeygi and A. Maleki. Geodesic k-means clustering. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, Dec. 2008.
6. P. Baker, C. Fermüller, Y. Aloimonos, and R. Pless. A spherical eye from multiple cameras (makes better models of the world). In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, page 576, Los Alamitos, CA, USA, 2001. IEEE Computer Society.

metric	Hessian	Gradient
$d_\angle(\mathtt{R},\mathtt{S}) = \theta$	$(1/2)\ \cot(\theta/2)(\mathtt{I} - \hat{\mathbf{w}}\hat{\mathbf{w}}^\top)$	$-\hat{\mathbf{w}}$
$d_{\text{chord}}(\mathtt{R},\mathtt{S}) = 2\sqrt{2}\ \sin(\theta/2)$	$\sqrt{2}/2\ \left(-\sin(\theta/2)\hat{\mathbf{w}}\hat{\mathbf{w}}^\top + \cos(\theta/2)\cot(\theta/2)(\mathtt{I} - \hat{\mathbf{w}}\hat{\mathbf{w}}^\top)\right)$	$-\sqrt{2}\ \cos(\theta/2)\ \hat{\mathbf{w}}$
$d_{\text{quat}}(\mathtt{R},\mathtt{S}) = 2\sin(\theta/4)$	$1/8\ \left(-\sin(\theta/4)\ \hat{\mathbf{w}}\hat{\mathbf{w}}^\top + \cos(\theta/2)/\sin(\theta/4)(\mathtt{I} - \hat{\mathbf{w}}\hat{\mathbf{w}}^\top)\right)$	$-(1/2)\ \cos(\theta/4)\hat{\mathbf{w}}$
$d_\angle(\mathtt{R},\mathtt{S})^2 = \theta^2$	$2\,\hat{\mathbf{w}}\hat{\mathbf{w}}^\top + \theta\cot(\theta/2)\,(\mathtt{I} - \hat{\mathbf{w}}\hat{\mathbf{w}}^\top)$	$-2\theta\hat{\mathbf{w}}$
$d_{\text{chord}}(\mathtt{R},\mathtt{S})^2 = 8\sin^2(\theta/2)$	$4\left(\cos\theta\,\hat{\mathbf{w}}\hat{\mathbf{w}}^\top + \cos^2(\theta/2)(\mathtt{I} - \hat{\mathbf{w}}\hat{\mathbf{w}}^\top)\right)$	$-4\sin(\theta)\hat{\mathbf{w}}$
$d_{\text{quat}}(\mathtt{R},\mathtt{S})^2 = 4\sin^2(\theta/4)$	$(1/2)\ \cos(\theta/2)\ \mathtt{I}$	$-\sin(\theta/2)\hat{\mathbf{w}}$

Table 3 Hessians and gradient of the different distance metrics $f(\mathtt{R}) = d^p(\mathtt{R},\mathtt{S})$, expressed in terms of the coordinate system induced by the exponential map $\exp_{\mathtt{R}}$ at $\mathtt{R}$. Here $\theta\hat{\mathbf{w}}$ is the angle-axis representation of $\mathtt{R}^\top\mathtt{S}$, namely $\theta\hat{\mathbf{w}} = \log(\mathtt{R}^\top\mathtt{S})$.

metric	$g'(\cos(\theta/2))$	$g''(\cos(\theta/2))$
$d_\angle(\mathtt{R},\mathtt{L}) = 2\arccos(C)$	$-2/\sin(\theta/2)$	$-2\cos(\theta/2)/\sin^3(\theta/2)$
$d_{\text{chord}}(\mathtt{R},\mathtt{L}) = \sqrt{8(1-C^2)}$	$-2\sqrt{2}\ \cot(\theta/2)$	$-2\sqrt{2}/\sin^3(\theta/2)$
$d_{\text{quat}}(\mathtt{R},\mathtt{L}) = \sqrt{2(1-C)}$	$-1/(2\sin(\theta/4))$	$-1/(8\sin^3(\theta/4))$
$d_\angle(\mathtt{R},\mathtt{L})^2 = 4\arccos(C)^2$	$-4\theta/\sin(\theta/2)$	$-4(\theta\cos(\theta/2) - 2\sin(\theta/2))/\sin^3(\theta/2)$
$d_{\text{chord}}(\mathtt{R},\mathtt{L})^2 = 8(1-C^2)$	$-16\cos(\theta/2)$	-16
$d_{\text{quat}}(\mathtt{R},\mathtt{L})^2 = 2(1-C)$	-2	0

Table 4 Hessians and gradient of the conjugate cost function $f(\mathtt{S}) = d^p(\mathtt{RS},\mathtt{SL})$, evaluated at $\mathtt{S} = \mathtt{I}$. The gradient is $g'\,\nabla_C$ and the Hessian is $g'\mathtt{H}_C + g''\,\nabla_C\,\nabla_C{}^\top$, where ∇_C and $\mathtt{H}_C$ are given by (32) and (33).

7. E. Beltrami. Teoria fondamentale degli spazii di curvatura costante. *Annali di Matematica pura ed applicata*, II(2nd series):232–255, 1868.
8. S. Buchholz and G. Sommer. On averaging in Clifford groups. *Computer Algebra and Geometric Algebra with Applications*, pages 229–238, 2005.
9. É. Cartan. *Leçons sur la géométrie des espaces de Riemann*. Gauthier-Villars, 2nd edition, 1951.
10. B. Clipp, J.-H. Kim, J.-M. Frahm, M. Pollefeys, and R. Hartley. Robust 6DOF motion estimation for non-overlapping multi-camera systems. In *Workshop on Applications of Computer Vision, WACV08*, pages 1 – 8, January 2008.
11. J. M. Corcuera and W. S. Kendall. Riemannian barycentres and geodesic convexity. *Math. Proc. Camb. Phil. Soc.*, 127:253–269, 1999.
12. Y. Dai, J. Trumpf, H. Li, N. Barnes, and R. Hartley. Rotation averaging with application to camera-rig calibration. In *Proc. Asian Conference on Computer Vision, Xian*, 2009.
13. K. Daniilidis. Hand-eye calibration using dual quaternions. *International Journal of Robotics Research*, 18:286–298, 1998.
14. D. Devarajan and R. J. Radke. Calibrating distributed camera networks using belief propagation. *EURASIP J. Appl. Signal Process.*, 2007(1), 2007.
15. U. Eckhardt. Weber's problem and Weiszfeld's algorithm in general spaces. *Mathematical Programming*, 18(1):186–196, 1980.
16. A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.
17. S. Esquivel, F. Woelk, and R. Koch. Calibration of a multi-camera rig from non-overlapping views. In *In DAGM07*, pages 82–91, 2007.
18. S. Fiori and T. Tanaka. An averaging method for a committee of special-orthogonal-group machines. In *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, pages 2170–2173, May 2008.
19. P. Fletcher, C. Lu, and S. Joshi. Statistics of shape via principal geodesic analysis on lie groups. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–95–I–101 vol.1, June 2003.
20. P. T. Fletcher, S. Venkatasubramanian, and S. Joshi. The geometric median on Riemannian manifolds with applications to robust atlas estimation. *Neuroimage*, 45 (1 Suppl):143–152, 2009.
21. C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society, B*, 53(2):285 – 339, 1991.
22. V. M. Govindu. Combining two-view constraints for motion estimation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 218–225, Los Alamitos, CA , USA, 2001. IEEE Computer Society.
23. V. M. Govindu. Lie-algebraic averaging for globally consistent motion estimation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 684–691, Los Alamitos, CA , USA, 2004. IEEE Computer Society.
24. V. M. Govindu. Robustness in motion averaging. In *Proc. Asian Conference on Computer Vision*, pages 457–466, 2006.
25. C. Gramkow. On averaging rotations. *Int. J. Comput. Vision*, 42(1-2):7–16, 2001.
26. K. Grove, H. Karcher, and E. A. Ruh. Jacobi fields and Finsler metrics on compact Lie groups with an application to differentiable pinching problems. *Math. Ann.*, 211:7–21, 1974.
27. R. Hartley, K. Aftab, and J. Trumpf. L1-rotation averaging using the Weiszfeld algorithm. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2011.
28. R. Hartley and F. Kahl. Global optimization through rotation space search. *International Journal of Computer Vision*, 82(1):64 – 79, April 2009.
29. R. Hartley and F. Schaffalitzky. L_∞ minimization in geometric reconstruction problems. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Washington DC*, pages I–504–509, June 2004.
30. R. Hartley and J. Trumpf. A characterization of weakly convex sets in projective space. Preprint, 2009.

31. R. Hartley, J. Trumpf, and Y. Dai. Rotation averaging and weak convexity. In *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, pages 2435–2442, 2010.
32. R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision – 2nd Edition*. Cambridge University Press, 2004.
33. B. K. P. Horn, H. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America*, 5(7):1127–1135, 1988.
34. M. Humbert, N. Gey, J. Muller, and C. Esling. Determination of a mean orientation from a cloud of orientations. Application to electron back-scattering pattern measurements. *Journal of Applied Crystallography*, 29(6):662–666, Dec 1996.
35. M. Humbert, N. Gey, J. Muller, and C. Esling. Response to Morawiec's (1998) comment on *Determination of a mean orientation from a cloud of orientations. Application to electron back-scattering pattern measurements* . *Journal of Applied Crystallography*, 31(3):485, Jun 1998.
36. K. Hüper. A calculus approach to matrix eigenvalue algorithms. Habilitationsschrift, Universität Würzburg, Germany, July 2002.
37. F. Kahl. Multiple view geometry and the L_∞-norm. In *Proc. International Conference on Computer Vision*, pages 1002–1009, 2005.
38. F. Kahl and R. Hartley. Multiple view geometry under the L_∞-norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9):1603 – 1617, September 2008.
39. K. Kanatani. *Group-theoretical methods in image understanding*. Springer, 1990.
40. H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30(5):509–541, 1977.
41. R. Kaucic, R. I. Hartley, and N. Y. Dano. Plane-based projective reconstruction. In *Proc. 8th International Conference on Computer Vision, Vancouver, Canada*, pages I–420–427, 2001.
42. J.-H. Kim, R. Hartley, J.-M. Frahm, and M. Pollefeys. Visual odometry for non-overlapping views using second-order cone programming. In *Proc. Asian Conference on Computer Vision*, volume 2, pages 353 – 362, November 2007.
43. J.-H. Kim, H. Li, and R. Hartley. Motion estimation for multi-camera systems using global optimization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
44. J.-H. Kim, H. Li, and R. Hartley. Motion estimation for non-overlapping multi-camera rigs: Linear algebraic and L_∞ geometric solutions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1044 – 1059, June 2010.
45. K. Krakowski, K. Hüper, and J. Manton. On the computation of the Karcher mean on spheres and special orthogonal groups. In *RoboMat 2007, Workshop on Robotics and Mathematics*, Coimbra, Portugal, 2007.
46. R. Kumar, A. Ilie, J.-M. Frahm, and M. Pollefeys. Simple calibration of non-overlapping cameras with a mirror. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7, June 2008.
47. H. Le. Locating Fréchet means with application to shape spaces. *Advances in Applied Probability*, 33:324–338, 2001.
48. H. Le. Estimation of Riemannian barycentres. *LMS Journal of Computation and Mathematics*, 7:193–200, 2004.
49. P. Lébraly, C. Deymier, O. Ait-Aider, E. Royer, and M. Dhome. Flexible extrinsic calibration of non-overlapping cameras using a planar mirror: Application to vision-based robotics. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 5640–5647. IEEE, 2010.
50. H. Li, R. Hartley, and J.-H. Kim. Linear approach to motion estimation using generalized camera models. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
51. Y. Li. A Newton acceleration of the Weiszfeld algorithm for minimizing the sum of euclidean distances. *Computational Optimization and Applications*, 10:219–242, 1998.
52. F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, 4(4):333–349, 1997.
53. J. H. Manton. A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups. In *Proceedings of the Eighth International Conference on Control, Automation, Robotics and Vision*, pages 2211–2216, Kunming, China, December 2004.
54. F. Markley, Y. Cheng, J. Crassidis, and Y. Oshman. Averaging quaternions. *Journal of Guidance, Control, and Dynamics*, 30(4):1193–1197, 2007.
55. D. Martinec and T. Pajdla. Robust rotation and translation estimation in multiview reconstruction. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.
56. W. Massey. *Algebraic topology: An introduction*. Springer, 1977.
57. M. Moakher. Means and averaging in the group of rotations. *SIAM J. Matrix Anal. Appl.*, 24(1):1–16 (electronic), 2002.
58. A. Morawiec. Comment on *Determination of a mean orientation from a cloud of orientations. Application to electron back-scattering pattern measurements by Humbert et al (1996)* . *Journal of Applied Crystallography*, 31(3):484, Jun 1998.
59. A. Morawiec. A note on mean orientation. *Journal of Applied Crystallography*, 31(5):818–819, Oct 1998.
60. A. Morawiec. *Orientations and Rotations: Computations in Crystallographic Textures*. Springer, 2004.
61. S. Myers. Arcs and geodesics in metric spaces. *Transactions of the American Mathematical Society*, 57(2):217–227, 1945.
62. J. Nocedal and S. Wright. *Numerical optimization*. Springer, 1999.
63. L. Ostresh. Convergence of a class of iterative methods for solving weber location problem. *Operations Research*, 26:597–609, 1978.
64. F. Park and B. Martin. Robot sensor calibration: solving AX=XB on the euclidean group. *Robotics and Automation, IEEE Transactions on*, 10(5):717–721, Oct 1994.
65. X. Pennec. Computing the mean of geometric features: Application to the mean rotation. Technical Report INRIA RR-3371, INRIA, 1998.
66. R. Pless. Using many cameras as one. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 587–593, 2003.
67. C. Qi, K. A. Gallivan, and P.-A. Absil. Riemannian BFGS algorithm with applications. In M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels, editors, *Recent Advances in Optimization and its Applications in Engineering*, pages 183–192. Springer, 2010.
68. B. Rinner and W. Wolf. A bright future for distributed smart cameras. *Processings of the IEEE*, 96(10):1562–1564, 2008.
69. R. Rockafellar. *Convex analysis*. Princeton University Press, 1970.
70. R. Rodrigues, J. Barreto, and U. Nunes. Camera pose estimation using images of planar mirror reflections. *Computer Vision–ECCV 2010*, pages 382–395, 2010.
71. C. Rother and S. Carlsson. Linear multi view reconstruction and camera recovery. In *Proc. 8th International Conference on Computer Vision, Vancouver, Canada*, pages I–42–49, 2001.
72. A. Sarlette and R. Sepulchre. Consensus optimization on manifolds. *SIAM J. Control Optim.*, 48(1):56–76, 2009.
73. K. Sim and R. Hartley. Recovering camera motion using L_∞ minimization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, New York City*, pages 1230 – 1237, 2006.
74. J. Steiner. Einige Gesetze über die Theilung der Ebene und des Raumes. *J. Reine Angew. Math.*, 1:349–364, 1826.
75. K. Strobl and G. Hirzinger. Optimal hand-eye calibration. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 4647–4653, Oct. 2006.

76. P. Sturm and T. Bonfort. How to compute the pose of an object without a direct view? *Computer Vision–ACCV 2006*, pages 21–31, 2006.
77. R. Subbarao and P. Meer. Nonlinear mean shift over Riemannian manifolds. *Int. J. Comput. Vision*, 84(1):1–20, 2009.
78. S. Teller, M. Antone, Z. Bodnar, M. Bosse, S. Coorg, M. Jethwa, and N. Master. Calibrated, registered images of an extended urban area. *Int. J. Comput. Vision*, 53(1):93–107, 2003.
79. R. Tron, R. Vidal, and A. Terzis. Distributed pose averaging in camera networks via consensus on SE(3). In *Second ACM/IEEE International Conference on Distributed Smart Cameras*, Sept. 2008.
80. A. Weber. *Über den Standort der Industrien. Teil 1, Reine Theorie des Standorts*. Tübingen: J.C.B. Mohr, 1909.
81. E. Weiszfeld. Sur le point pour lequel la somme des distances de n points donnes est minimum. *Tohoku Math. Journal*, 43:355–386, 1937.
82. F. Wu, Z. Wang, and Z. Hu. Cayley transformation and numerical stability of calibration equation. *International Journal of Computer Vision*, 82(2):156–184, 2009.
83. L. Yang. Riemannian median and its estimation. *LMS Journal of Computation and Mathematics*, 13:461–479, 2010.
84. H. Zhang. Hand/eye calibration for electronic assembly robots. *Robotics and Automation, IEEE Transactions on*, 14(4):612–616, Aug 1998.