A GLOBALLY CONVERGENT NUMERICAL ALGORITHM FOR COMPUTING THE CENTRE OF MASS ON COMPACT LIE GROUPS

Jonathan H. Manton

Department of Electrical and Electronic Engineering
The University of Melbourne, Victoria 3010, Australia.

ABSTRACT

Motivated by applications in fuzzy control, robotics and vision, this paper considers the problem of computing the centre of mass (precisely, the Karcher mean) of a set of points defined on a compact Lie group, such as the special orthogonal group consisting of all orthogonal matrices with unit determinant. An iterative algorithm, whose derivation is based on the geometry of the problem, is proposed. It is proved to be globally convergent. Interestingly, the proof starts by showing the algorithm is actually a Riemannian gradient descent algorithm with fixed step size.

1. INTRODUCTION

A fundamental operation in signal processing and control is averaging. For instance, given a set of noise corrupted measurements, the signal to noise ratio is improved by averaging. Geometrically, the measurements can be thought of as points scattered around the true value and averaging computes the centre of mass of these points. If the points don't lie in a flat space like Euclidean space though, but instead lie on a curved space like the sphere — such as geographical locations marked on a globe of the world do — then the averaging operator is generally more complicated.

On the sphere, or more generally any Riemannian manifold M, one averaging operator is the Karcher mean [8] of the points. Specifically, if $d(\cdot,\cdot)$ is the distance function on M, the Karcher mean of the points $Q_1, \cdots, Q_k \in M$ is the point X minimising

$$f(X) = \frac{1}{2k}\sum_{i=1}^{k} d^2(Q_i, X). \qquad (1)$$

This reduces to the usual averaging operator when M is a Euclidean space. If M has positive curvature though (such as if M is a sphere), there might not be a unique minimum of (1). It is known though that as long as the points $Q_1, \cdots, Q_k$ are not spaced too far apart, the global minimum of (1) is unique [8, 9].

This work was supported by the Australian Research Council. Presented at ICARCV 2004, Kunming, China. Email: j.manton@ee.mu.oz.au

The Karcher mean of points on a sphere has been studied, amongst other places, in [5], where it is shown the sphere is relatively special. Indeed, the sphere is the only compact manifold for which the Euclidean concept of centre of mass extends. (In Euclidean space, although the centre of mass and the Karcher mean correspond to the same point, their definitions are different.)

Another well-known non-negatively curved and compact space is the special orthogonal group

$$\mathcal{SO}(n) = \left\{X \in \mathbb{R}^{n\times n} :\ X^T X = I,\ \det X = 1\right\} \qquad (2)$$

consisting of all orthogonal matrices with unit determinant. (Every orthogonal matrix has a determinant of either 1 or -1, so the special orthogonal group is the connected component containing the identity matrix of the orthogonal group.) The case $n = 3$ is especially well-known, since $\mathcal{SO}(3)$ is the set of rotations acting on $\mathbb{R}^3$. It has been studied in [13], and applications requiring finding the Karcher mean include [1, 10, 12].

Despite the Karcher mean being a fundamental operation, the author is unaware of any general purpose algorithm for computing the Karcher mean. For instance, no numerical algorithm for $\mathcal{SO}(3)$, much less $\mathcal{SO}(n)$, was given or referenced in the recent study [13]. The contribution of the present paper is to derive such a general purpose algorithm for a special class of manifolds, namely compact Lie groups (which include the special orthogonal group), and to prove global convergence of said algorithm. Here, global convergence is under the necessary proviso that the points $Q_1, \cdots, Q_k$ are sufficiently close together; recall that if this is not so, the Karcher mean is not well defined.

The reason for considering only compact Lie groups is twofold. Firstly, the proposed algorithm exploits the natural correspondence between a Lie group and its associated Lie algebra. The algorithm also exploits the fact that compact Lie groups, as opposed to more general Lie groups, have non-negative curvature. Extending the algorithm to more general Lie groups and other Riemannian manifolds will be the topic of future study.

The reader is referred to [2, 4, 14] for general information on Lie groups and Riemannian manifolds.

2. LIE GROUPS AND THE ALGORITHM

For the impatient reader, the proposed algorithm is stated up front. In Algorithm 1, $\mathcal{G}$ denotes a compact Lie group and $\mathfrak{g}$ its corresponding Lie algebra. Recall that the Karcher mean is defined to be the minimising X of (1).

If $\mathcal{G} = \mathcal{SO}(n)$ then $\exp$ and $\log$ are the matrix exponential and principal matrix logarithm functions respectively. Moreover, the standard Frobenius norm can be used in Step 3 of Algorithm 1 in this case.

Algorithm 1 *Given points $Q_1, \cdots, Q_k \in \mathcal{G}$, compute the Karcher mean $X \in \mathcal{G}$.*

1. *Set $X := Q_1$. Choose desired tolerance $\epsilon > 0$.*
2. *Compute $A := \frac{1}{k}\sum_{i=1}^{k} \log(X^{-1}Q_i)$.*
3. *If $\|A\| < \epsilon$ then stop.*
4. *Update $X := X\exp(A)$ and goto step 2.*

Notes: i) $\exp$ denotes the Lie group exponential map from $\mathcal{G}$ to $\mathfrak{g}$. Its inverse about the origin is denoted by $\log$. It is implicitly assumed $Q_1, \cdots, Q_k$ are sufficiently close together that $X^{-1}Q_i$ remains in the domain of definition of $\log$. See Section 6 for conditions under which this is true. ii) Any norm on the Lie algebra $\mathfrak{g}$ can be used in step 3. See Section 6 for more information. iii) In practice, numerical errors can cause the iterate X to diverge from the manifold $\mathcal{G}$. This can be prevented by projecting the iterate back onto the manifold after each iteration. For example, if $X = USV^T$ is the singular value decomposition of a matrix $X \in \mathbb{R}^{n\times n}$, then $\pi(X)$, the Euclidean projection of X onto $\mathcal{SO}(n)$, is given by $\pi(X) = UV^T$; see [11] for a proof. Alternatively, projecting A onto the Lie algebra straight after step 2 appears to work well and is usually numerically cheaper to compute. For $\mathcal{SO}(n)$, this is achieved by setting $A := (A - A^T)/2$.

An example of the performance of the algorithm appears in the next section. The required concepts from Lie group theory are now introduced.

2.1. A Metric Structure for Lie Groups

A Lie group $\mathcal{G}$ is simultaneously a smooth differentiable manifold and a group, and moreover, the group operations of multiplication and inversion are smooth mappings [3, p. 81]. The example used throughout this paper of a (compact, connected, semisimple) Lie group is $\mathcal{SO}(n)$, defined in (2). Multiplication and inversion on $\mathcal{SO}(n)$ are given by standard matrix multiplication and inversion.

The distinguishing feature of a Lie group compared to other manifolds is its symmetry. Perhaps the simplest (but non-compact) Lie group is $\mathbb{R}^n$ with group multiplication given by vector addition. In $\mathbb{R}^n$, every point looks like every other point, in that any neighbourhood $N \subset \mathbb{R}^n$ of a point $x \in \mathbb{R}^n$ can be translated to a neighbourhood $N - x$ about the origin without altering the geometry.

The same is true on an arbitrary Lie group $\mathcal{G}$, but due to the non-commutativity, there are two different ways of translating points, namely left translation and right translation. Left translation is the operation $l_Y : \mathcal{G} \to \mathcal{G}$ mapping X to YX, where $Y \in \mathcal{G}$. Similarly, right translation r_Z maps X to XZ.

If the Karcher mean is to reflect the Lie group structure then the Karcher mean must be left and right translation invariant, meaning that for all $Y \in \mathcal{G}$, the Karcher mean of $l_Y(Q_1), \cdots, l_Y(Q_k)$ is $l_Y(X)$, where X is the Karcher mean of $Q_1, \cdots, Q_k$, and similarly for right translation. This is so if the distance function in (1) is itself bi-invariant.

There is a standard method of making a compact semisimple Lie group $\mathcal{G}$ into a Riemannian manifold in such a way that the resulting distance function is bi-invariant. The method can be extended to an arbitrary compact Lie group. This is now explained.

Let $\mathcal{G}$ be a compact Lie group. Since it is a manifold, it has a tangent space $T_X\mathcal{G}$ associated with it about each point $X \in \mathcal{G}$. For example, thinking of $\mathcal{SO}(n)$ as a submanifold of $\mathbb{R}^{n\times n}$, the tangent space $T_X\mathcal{SO}(n)$ at $X \in \mathcal{SO}(n)$ is identified with the vector space

$$T_X\mathcal{SO}(n) = \left\{XA : \ A \in \mathbb{R}^{n\times n}, \ A^T = -A\right\}. \qquad (3)$$

Let I denote the identity element of $\mathcal{G}$. (If $\mathcal{G} = \mathcal{SO}(n)$ then I is simply the identity matrix.) The tangent space at I is called the Lie algebra of $\mathcal{G}$ and denoted by $\mathfrak{g}$. If $\mathcal{G} = \mathcal{SO}(n)$ then $\mathfrak{g} = \mathfrak{so}(n)$ where $\mathfrak{so}(n)$ is the space of $n \times n$ skew-symmetric matrices. Note the dimension of $\mathfrak{so}(n)$, and hence of $\mathcal{SO}(n)$, is $n(n-1)/2$.

Differentiating the left translation operator l_Y yields the linear map $(l_Y)_* : T_X\mathcal{G} \to T_{YX}\mathcal{G}$ sending U to YU. Similarly, $(r_Z)_* : T_X\mathcal{G} \to T_{ZX}\mathcal{G}$ sends U to UZ. Given a vector $A \in \mathfrak{g} = T_I\mathcal{G}$, it can be pushed forward using $(l_Y)_*$ to a vector in the tangent space about the point $Y \in \mathcal{G}$. The resulting vector field on $\mathcal{G}$ is called a left invariant vector field.

The Lie bracket of two vector fields on a manifold is an operation returning a vector field [15, p. 36]. When applied to two left invariant vector fields on a Lie group, it returns a left invariant vector field. Since a left invariant vector field is specified uniquely by its value at the identity, the Lie bracket induces an operation on the Lie algebra. On $\mathcal{SO}(n)$, the Lie bracket of $A, B \in \mathfrak{so}(n)$ is $[A, B] = AB - BA$.

Define the linear map $\mathrm{ad}_A : \mathfrak{g} \to \mathfrak{g}$ by $\mathrm{ad}_A\, B = [A, B]$. The Killing form of $A, B \in \mathfrak{g}$, denoted $\langle A, B\rangle$, is defined to be

$$\langle A, B\rangle = \mathrm{tr}\left\{\mathrm{ad}_A \circ \mathrm{ad}_B\right\}. \qquad (4)$$

On $\mathfrak{so}(n)$, $\langle A, B\rangle = -(n-2)\,\mathrm{tr}\left\{B^T A\right\}$. In this case, and indeed for any compact semisimple Lie group, the Killing

form is negative definite. The Lie algebra of a compact Lie group decomposes [7, Proposition 6.6] as a direct sum of its centre and the derived group $[\mathfrak{g}, \mathfrak{g}]$. The latter is semisimple. Therefore, after choosing an arbitrary negative definite bilinear form on the centre of $\mathfrak{g}$, a negative definite bilinear form can be defined on the whole of $\mathfrak{g}$ by using the chosen form on the centre and the Killing form on the derived group.

Given a negative definite bilinear form on $\mathfrak{g}$, the inner product of two tangent vectors $U, V \in T_X\mathcal{G}$ can be defined by

$$\langle U, V\rangle_X = -c\,\langle (l_{X^{-1}})_* U, (l_{X^{-1}})_* V\rangle \tag{5}$$

where $c > 0$ is an arbitrary constant. Alternatively, right translation could have been used. The key feature of the above choice of bilinear form involving the Killing form is that left and right translation lead to the same inner product, or in other words, $\langle U, V\rangle_X$ is a bi-invariant metric. This means the induced distance function is bi-invariant, as required. Note that on $\mathcal{SO}(n)$, the resulting metric structure is

$$\langle U, V\rangle_X = \operatorname{tr}\left\{V^T U\right\}, \quad U, V \in T_X\mathcal{SO}(n). \tag{6}$$

2.2. Geodesics and the Distance Function

In order to determine the distance function, it is necessary to know what the curve of shortest distance between two points looks like. Such curves are called geodesics, and on compact manifolds, existence of a geodesic of shortest length connecting any two points is assured. This leads to the introduction of the Lie group exponential map.

A one-parameter subgroup is a smooth curve $\gamma : \mathbb{R} \to \mathcal{G}$ satisfying: $\gamma(0) = I$, $\gamma(s + t) = \gamma(s)\gamma(t)$ and $\gamma(-t) = [\gamma(t)]^{-1}$. Note $\gamma'(0)$ is an element of $\mathfrak{g}$. In fact, the map from a one-parameter subgroup γ to an element $\gamma'(0)$ of the Lie algebra is a bijection. The inverse of this map is called the Lie group exponential map and is denoted by $\exp : \mathcal{G} \to \mathfrak{g}$. That is, if $A \in \mathfrak{g}$ then $\gamma(t) = \exp(At)$ is the unique one-parameter subgroup with $\gamma'(0) = A$. If $\mathcal{G} = \mathcal{SO}(n)$ then $\exp$ is the standard matrix exponential map.

It is a standard result that the geodesics of a Lie group $\mathcal{G}$ equipped with a bi-invariant metric (5) are the curves $\gamma(t) = X\exp(At)$, where $X \in \mathcal{SO}(n)$ and $A \in \mathfrak{so}(n)$ are arbitrary. It follows that the distance between the points X and $X\exp(A)$, provided they are not so far apart that there is a shorter geodesic connecting them, is $\sqrt{\langle A, A\rangle} = \sqrt{\operatorname{tr}\{-A^2\}}$. Thus, the squared distance between any two points is

$$d^2(X, Y) = \min_{\substack{A \\ \exp(A)=X^T Y}} \langle A, A\rangle. \tag{7}$$

Let $\log$ denote the inverse of $\exp$ about the origin. Then, provided X and Y are sufficiently close,

$$d^2(X, Y) = \left\langle \log(X^T Y), \log(X^T Y)\right\rangle. \tag{8}$$

On $\mathcal{SO}(n)$, $d^2(X, Y) = (-1/2)\operatorname{tr}\left\{\left[\log(X^T Y)\right]^2\right\}$, with $\log$ denoting the standard principal logarithm of a matrix. How close X and Y need to be for (8) to be valid is discussed in Section 4. (On $\mathcal{SO}(n)$, the above formula is valid globally.)

2.3. Derivation of The Algorithm

Algorithm 1 was derived as follows. A necessary condition for X to be the Karcher mean is for the gradient of (1) to be zero.

Lemma 2 *The gradient of (1) is*

$$\operatorname{grad} f(X) = -\frac{1}{k} X \sum_{i=1}^{k} \log(X^{-1} Q_i). \tag{9}$$

PROOF. Let $\exp_X : T_X\mathcal{G} \to \mathcal{G}$ denote the Riemannian exponential map. Its relation to the Lie group exponential map is $\exp_X(XA) = X\exp(A)$. Theorem 1.2 of [8] implies $\operatorname{grad} f(X) = (-1/k)\sum_{i=1}^{k} \exp_X^{-1} Q_i$. The lemma follows. □

It is straightforward to deduce that $\operatorname{grad} f(X) = 0$ if and only if $\frac{1}{k}\sum_{i=1}^{k} \log(X^{-1} Q_i) = 0$. Geometrically, this condition says that the ordinary centre of mass of the points $\log(X^{-1} Q_i)$ for $i = 1, \cdots, k$ is at the origin of the Lie algebra.

The second observation is that the Lie algebra serves as the first order approximation of the Lie group about the identity. Therefore, if the Q_i are very close to the identity, one would expect the Karcher mean to be approximately $\exp\left(\frac{1}{k}\sum_{i=1}^{k} \log(Q_i)\right)$.

These two observations motivated Algorithm 1. However, they do not on their own guarantee that Algorithm 1 actually converges. After presenting a numerical example, the rest of this paper is concerned with proving Algorithm 1 is a globally convergent algorithm provided the Q_i are all within a certain distance of each other.

3. A NUMERICAL EXAMPLE

Three points $Q_1, Q_2, Q_3 \in \mathcal{SO}(5)$ were generated at random. Since $\mathcal{SO}(5)$ is a ten dimensional space, these points are plotted as asterisks in Figure 1 by first mapping them to the respective points $\log Q_1, \log Q_2, \log Q_3$ in the Lie algebra and then projecting onto a two dimensional subspace.

To illustrate the curvature of $\mathcal{SO}(5)$, the geodesic triangle formed by Q_1, Q_2, Q_3 is also plotted. In Euclidean space, the Karcher mean of three points can be found by first drawing the triangle connecting the three points and then drawing three lines, each from a vertex to the midpoint of the opposite side. These three lines will meet at a single point, the Karcher mean. In $\mathcal{SO}(5)$, a similar construction

can be undertaken, but in ten dimensions, these three lines will not meet. Even when projected onto a two dimensional subspace, Figure 1 shows these lines do not meet at a single point.

Algorithm 1 is used to compute the Karcher mean of the three points. The sequence of iterates is plotted in Figure 1 as circles. These circles appear to converge to the Karcher mean after the second iteration. In finer detail, Figure 2 shows the norm of the gradient of the cost function $f(X)$ decreasing at a linear rate until machine precision is reached on the 15th iterate.

4. CONVEXITY ON LIE GROUPS

The Lie group $\mathcal{SO}(2)$ is isomorphic to a circle, with distance measured by arc length. By considering two or more equally spaced points around the circle, it becomes clear that the global minimum of (1) is not unique in general.

Therefore, the result this paper aims for is that there exists a calculable radius r such that if $Q_1, \cdots, Q_i \in B(I, r)$, the open ball of radius r centred at the identity of $\mathcal{G}$, then the Karcher mean exists, is unique, and Algorithm 1 converges to it.

The first restriction on r comes from the need for $\log$ in (8) to be well-defined. The radius of injectivity is, by definition, the largest ρ such that $\exp$ is a diffeomorphism from $B(0, \rho)$ onto its image. Here, $B(0, \rho)$ denotes the open ball centred at the origin of $\mathfrak{g}$. In particular, r must be such that $B(I, r)$ is a subset of $\exp(B(0, \rho))$. However, $\exp(B(0, \rho)) = B(I, \rho)$, hence the requirement is simply $r \leq \rho$.

The second restriction is that $B(I, r)$ be a convex set and $g(X) = d(I, X)$ be a convex function on $B(I, r)$, and indeed, the largest such r is called the convexity radius. Here, a set $\Omega \subset \mathcal{G}$ is convex if for any $X, Y \in \mathcal{G}$ there is a unique geodesic wholly contained in Ω connecting X to Y and such that it is also the unique minimising geodesic in $\mathcal{G}$ connecting X to Y. (Such a set is called strongly convex in [4].) A function $f : \Omega \to \mathbb{R}$ is convex if for any geodesic $\gamma : [0, 1] \to \Omega$, the function $f \circ \gamma : [0, 1] \to \mathbb{R}$ is convex in the usual sense, that is,

$$f(\gamma(t)) \leq (1 - t)f(\gamma(0)) + tf(\gamma(1)), \quad t \in [0, 1]. \quad (10)$$

If strict inequality holds then f is strictly convex.

These two restrictions are quite standard for Karcher means [4, 8]. It is not necessary though to use the same metric to define the open balls above as the metric (6) used to define the cost function (1). In fact, larger open balls can be obtained by using "better" norms. This generalisation is not pursued in full here though for reasons of space.

The third restriction is that $f(X)$ in (1) be strictly convex. Indeed, the convergence proof for Algorithm 1 is based on showing it is actually a Riemannian gradient descent algorithm applied to a strictly convex function.

5. JACOBI FIELDS AND BOUNDS

The previous section explained why there are restrictions on how far apart the points can be if the Karcher mean is to be unique and Algorithm 1 convergent. This section gives closed form expressions for these restrictions. It is based closely on the ideas and results in [6, 8].

Define on $\mathfrak{g}$ the norm $\|A\|^2 = \langle A, A \rangle$, where the inner product is the bi-invariant inner product constructed according to the procedure in Section 2.1. By scaling the inner product if necessary, it is assumed — call this assumption **(A1)** — that the chosen inner product induces a norm satisfying the inequality

$$\|[A, B]\| \leq \|A\| \cdot \|B\|, \quad A, B \in \mathfrak{g}. \quad (11)$$

For instance, it can be shown (6) (evaluated at $X = I$) induces a norm satisfying (11) if $\mathfrak{g} = \mathfrak{so}(n)$.

Recall that the construction in Section 2.1 required an arbitrary choice of an inner product on the centre of $\mathfrak{g}$. Because this choice does not affect the inequality (11), it is necessary to make the further assumption **(A2)** that the injectivity radius of $\exp$ restricted to the centre is at least π. This is always achievable by scaling the chosen inner product on the centre of $\mathfrak{g}$.

Under (A1) and (A2), Proposition 1.4 of [6] proves the injectivity radius of $\exp$ is at least π.

Under (A1) and (A2), Theorem 3.5 of [6] proves the convexity radius is at least $\pi/2$.

To enforce the third requirement that $f(X)$ be strictly convex, bounds on the eigenvalues of the Riemannian Hessian of $f(X)$ must be determined. As shown in [8], the Hessian of $f(X)$ can be expressed in terms of Jacobi fields. Moreover, bounds on the Hessian are obtained by using standard bounds on how much the curvature of the space can affect these Jacobi fields. It follows immediately from the definition of sectional curvature (see [2]) that, under (A1), the sectional curvatures of $\mathcal{G}$ lie in the interval $[0, 1/4]$. This knowledge is enough to apply the results of [8] to obtain the required bounds. However, as the first step towards generalising the results here to "better" norms (see Section 4), it is shown how the results of [6] and [8] can be combined.

The Hessian of $f(X)$ about the point X is found as follows [8]. Let $\gamma : (-\epsilon, \epsilon) \to \mathcal{G}$ denote a geodesic centred at X, that is, $\gamma(0) = X$. Let $c_i(s, t)$ be a family of geodesics from Q_i to $\gamma(t)$, with $c_i(0, t) = Q_i$ and $c_i(1, t) = \gamma(t)$. This family induces a Jacobi field along the geodesic $c_i(s, 0)$ from Q_i to X. Let $J_i(s)$ denote this Jacobi field and $J_i'(s)$ its covariant derivative. Note $J(0) = 0$ and

$J_i(1) = \dot{\gamma}(0)$. Then

$$\left.\frac{d^2}{dt^2}\right|_{t=0} f(\gamma(t)) = \frac{1}{k}\sum_{i=1}^{k} \langle J_i(1), J_i'(1)\rangle_X . \qquad (12)$$

Bounds on the eigenvalues of the Hessian are found by bounding $\langle J_i(1), J_i'(1)\rangle_X$. Since only norms and not inner products are considered in [6], the following readily verified formula is required:

$$\|A\|\left(\|A\| - \|B - A\|\right) \leq \langle A, B\rangle \leq \|A\|\|B\|. \qquad (13)$$

Lemma 3 *Let $r \in (0, \pi/2)$ and assume (A1) and (A2) hold. If $X, Q_1, \ldots, Q_k \in B(I, r)$ then the eigenvalues of the Hessian of (1) at X lie in the interval $[r/\tan(r), 1]$.*

PROOF. Define $J_i(s)$ as above. A geodesic from Q_i to X has length less than $2r$. It follows from [6, (2.3.2)] that $\|J_i'(0)\| \leq \|J_i(1)\|$. Since $J_i(1) = \dot{\gamma}(0)$, (12) and (13) imply $\left.\frac{d^2}{dt^2}\right|_{t=0} f(\gamma(t)) \leq \langle \dot{\gamma}(0), \dot{\gamma}(0)\rangle_X$, proving the largest eigenvalue is at most 1. Proposition 2.4 of [6] implies $\|J_i(1) - J_i'(1)\| \leq \left(1 - \frac{r}{\tan r}\right)\|J(1)\|$. Hence (12) and (13) imply $\left.\frac{d^2}{dt^2}\right|_{t=0} f(\gamma(t)) \geq \frac{r}{\tan r}\langle \dot{\gamma}(0), \dot{\gamma}(0)\rangle_X$. □

6. GLOBAL CONVERGENCE PROOF

The following lemma states that the Riemannian gradient descent algorithm with unit step size is guaranteed to converge if the eigenvalues of the Hessian are bounded away from zero and are at most one. It can be proved by extending the known proof in the Euclidean case to the Riemannian setting.

Lemma 4 *Let M be a Riemannian manifold and $f : M \to \mathbb{R}$ a function whose Riemannian Hessian has all its eigenvalues in the interval $[\delta, 1]$ for some $\delta > 0$. Let $\exp_x$ denote the Riemannian exponential map about the point $x \in M$. Then, for any $x_0 \in M$, the sequence*

$$x_{k+1} = \exp_{x_k}\left(-\operatorname{grad} f(x_k)\right) \qquad (14)$$

converges to the unique global minimum of f. Moreover, the distance from x to the minimum is bounded above by $\delta^{-1}\|\operatorname{grad} f(x)\|$.

Somewhat surprisingly, given the way it was derived, it turns out that Algorithm 1 is actually a Riemannian gradient descent algorithm.

Theorem 5 (Global Convergence) *Let $\mathcal{G}$ be a compact Lie group made into a Riemannian manifold by choosing a bi-invariant inner product on the Lie algebra (see Section 2.1). Moreover, assume the inner product is such that (A1) and (A2) of Section 5 hold. Assume $Q_1, \cdots, Q_k \in B(Y, r)$ for some $Y \in \mathcal{G}$ and $r \in (0, \pi/2)$. Then Algorithm 1 converges to the unique Karcher mean of $Q_1, \cdots, Q_k$. Moreover, if the norm in Step 3 of Algorithm 1 is the norm induced by the inner product on the Lie algebra then Algorithm 1 terminates only when X is within a distance of $\frac{\tan r}{r}\epsilon$ of the Karcher mean.*

PROOF. Note first that by exploiting the bi-invariance of the metric, it may be assumed without loss of generality that $Y = I$.

Sections 4 and 5 ensure the Karcher mean exists and is unique. In particular, Lemma 3 shows that $f(X)$ is strictly convex, implying it has a unique global minimum. (Existence and uniqueness of the Karcher mean is also proved in Theorem 3.7 of [6].)

Recall that the Riemannian exponential map $\exp_X$ on $\mathcal{G}$ is defined by $\exp_X(XA) = X\exp(A)$. Lemma 2 shows that $-\operatorname{grad} f(X) = XA$ where A is as in Step 2 of Algorithm 1. Thus, Step 4 of Algorithm 1 is equivalent to $X := \exp_X(XA) = \exp_X\left(-\operatorname{grad} f(X)\right)$. In particular, Algorithm 1 is indeed a Riemannian gradient descent algorithm with unit step size. The proof is completed by combining Lemma 4 with the bounds on the Hessian obtained in Lemma 3. □

It is remarked the bound $\frac{\tan r}{r}\epsilon$ in Theorem 5 is identical to the bound in [6, (3.7.1)]. The advantage of deriving it here based on Lemma 4 is that it emphasises the origin of the bound is merely the lower bound on the eigenvalues of the Hessian, and not something more sophisticated.

It is also remarked that the constraint $r < \pi/2$ in Theorem 5 is conservative. That is, for particular Lie groups, it might be possible to prove Algorithm 1 continues to converge for larger values of r.

Moreover, by exploiting further the ideas in [6], it is possible to replace the open ball $B(Y, r)$ in Theorem 5 by one with respect to a different norm. By an expedient choice of norm, the volume of the open ball can be significantly increased while still guaranteeing the convergence of Algorithm 1.

7. CONCLUSION

This paper proposed a general purpose algorithm for computing the Karcher mean of points on a compact Lie group. By showing the algorithm is actually a Riemannian gradient descent algorithm and then using Jacobi fields to compute bounds on the Hessian of the cost function (1), it was proved the algorithm is globally convergent provided the points are contained in an open ball of a particular size.

8. REFERENCES

[1] A. J. Abrantes and J. S. Marques. Shape tracking using centroid-based methods. In M. Figueiredo et al.,

editors, *Energy minimization methods in computer vision and pattern recognition*, Lect. Notes Comput. Sci. 2134, pages 576–591. Springer, 2001.

[2] A. Arvanitoyeorgos. *An Introduction to Lie Groups and the Geometry of Homogeneous Spaces*. American Mathematical Society, 2003.

[3] W. M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, second edition, 1986.

[4] I. Chavel. *Riemannian Geometry: A Modern Introduction*. Cambridge University Press, 1996.

[5] G. A. Gal'perin. A concept of the mass center of a system of material points in the constant curvature spaces. *Communications in Mathematical Physics*, 154:63–84, 1993.

[6] K. Grove, H. Karcher, and E. A. Ruh. Jacobi fields and Finsler metrics on compact Lie groups with an application to differentiable pinching problems. *Mathematische Annalen*, 211:7–21, 1974.

[7] S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*. American Mathematical Society, 2001.

[8] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977.

[9] W. S. Kendall. Probability, convexity, and harmonic maps with small image i: Uniqueness and fine existence. *Proceedings of the London Mathematical Society*, 61(3):371–406, 1990.

[10] F. Lewis, J. Campos, and R. Selmic. *Neuro-Fuzzy Control of Industrial Systems with Actuator Nonlinearities*. SIAM, 2002.

[11] J. H. Manton. Optimisation algorithms exploiting unitary constraints. *IEEE Transactions on Signal Processing*, 50(3):635–650, March 2002.

[12] J. S. Marques, J. M. Lemos, and A. J. Abrantes. A HMM approach to the estimation of random trajectories on manifolds. *Signal Processing*, 82(9):1205–1214, 2002.

[13] M. Moakher. Means and averaging in the group of rotations. *SIAM J. Matrix Anal. Appl.*, 24(1):1–16, 2002.

[14] P. Petersen. *Riemannian Geometry*. Springer-Verlag, 1998.

[15] F. W. Warner. *Foundations of Differentiable Manifolds and Lie Groups*. Springer-Verlag, 1983.

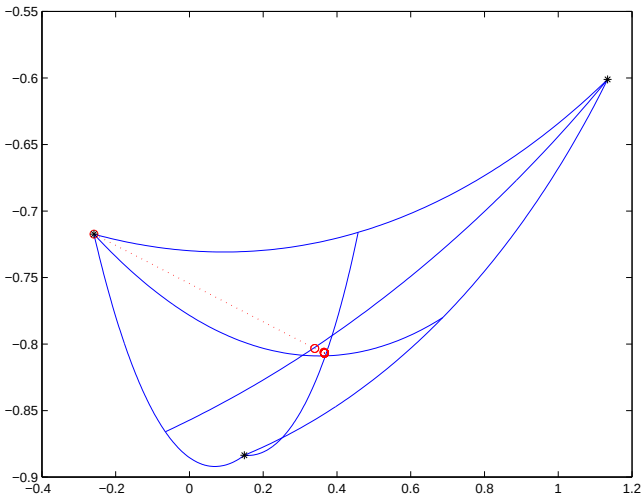

Figure 1: The three asterisks denote three points on $\mathcal{SO}(5)$. All lines are geodesics. The outer ones connect the three points together, forming a geodesic triangle. Each inner line connects a vertex to the midpoint of its opposing side. In Euclidean space, these inner lines always meet in a single point, the point being the Karcher mean. In curved spaces, as the figure shows, this is no longer true. The circles denote successive iterates of Algorithm 1. Although twenty circles are plotted, the circles have converged after the second iterate.

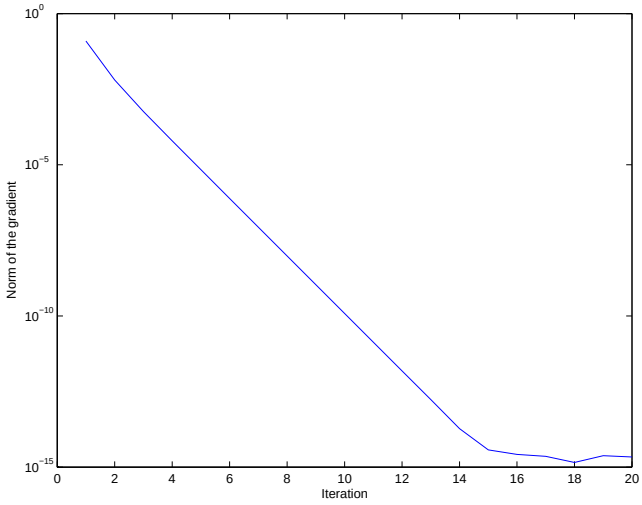

Figure 2: The norm of the gradient of the cost function (1) is plotted against the iteration number of Algorithm 1. Note the linear rate of convergence. Machine precision is reached after 15 iterations in this example.