On the Computation of the Karcher Mean on Spheres and Special Orthogonal Groups

Krzysztof A. Krakowski* Knut Hüper† Jonathan H. Manton‡

September 13, 2007

Abstract

This paper is concerned with computation of the *Karcher mean* on the unit sphere $\mathbf{S}^n$ and the special orthogonal group $\mathbb{SO}(n)$. The Karcher mean, or the *Riemannian centre of mass*, is defined as the point minimising the sum of the squared distances from that point to each of the given points. By its definition, the mean always belongs to the same space as the given points, however, it may not be unique. Motivated by applications in control, vision and robotics, this paper studies the numerical computation of the Karcher mean. We propose simpler and computationally more efficient gradient-like and Newton-like algorithms. We give explicit forms of these algorithms and show that if the set of points lie within a particular open ball, the algorithms are guaranteed to converge to the Karcher mean.

1 Introduction

The concept of an average, center of gravity or centroid of a set of points generalized to spaces other than $\mathbb{R}^n$ appears in the literature in a number of different contexts. Since the paper of Karcher [3] the average is often referred to as the Karcher mean, the term adopted here by the authors. This paper deals with a final set of points but the results extend to the case of a mass distribution.

Let $\mathbf{M}$ be a complete Riemannian manifold and Ω a final set of points in $\mathbf{M}$. Define a generalized *variance* to be the function $\Psi : \mathbf{M} \to \mathbb{R}$ given by

$$\Psi(x) := \frac{1}{2\#\Omega} \sum_{p \in \Omega} \operatorname{dist}(x, p)^2, \tag{1}$$

*School of Mathematics, Statistics and Computer Sciences, University of New England, Armidale, NSW 2350, Australia. E-mail:kris@mcs.une.edu.au.

†National ICT Australia, Canberra Research Laboratory, Locked Bag 8001, Canberra ACT 2601, Australia, and Department of Information Engineering, RSISE, The Australian National University, Canberra, ACT 0200, Australia. E-mail:knut.hueper@nicta.com.au.

‡Department of Information Engineering, Research School of Information Sciences and Engineering (RSISE) The Australian National University, Canberra, ACT 0200, Australia. E-mail:j.manton@ieee.org.

where $\mathrm{dist}(\cdot,\cdot)$ is the Riemannian (geodesic) distance in $\mathbf{M}$ and $\#\Omega$ is the cardinal, the number of points, of Ω. The *Karcher mean* $\bar{x} \in \mathbf{M}$ is a point where Ψ attains its minimum. It is known (*cf.* [3]) that if $\Omega \subset \mathscr{B}_\varrho$, where $\mathscr{B}_\varrho \subset \mathbf{M}$ is the open Riemannian ball of radius ϱ, and ϱ is sufficiently small so that $\mathscr{B}_\varrho$ is convex[1] and if $\varrho < \pi/4\,C^{-1/2}$, where C is a sectional curvature on $\mathscr{B}_\varrho$, then Ψ is convex on $\mathscr{B}_\varrho$. As a consequence Ψ has a unique point of local minimum in $\mathscr{B}_\varrho$, therefore the Karcher mean $\bar{x}$ is unique and belongs to $\mathscr{B}_\varrho$.

The motivation for this paper is twofold. The recent paper [6] considers $\mathbb{SO}(3)$, so it is natural to consider $\mathbb{SO}(n)$ for $n > 3$. In fact, $\mathbb{SO}(3)$ is a very special case because it has constant sectional curvature, whereas $\mathbb{SO}(n)$ doesn't, for $n > 3$. Therefore, not only do not the methods in [6] extend, it is not even clear in advance whether or not results valid for $n = 3$ remain true if $n > 3$. It is also remarked that no numerical algorithm for finding the Karcher mean was proposed in [6]. The main contribution of this paper are the new iterative algorithms whose convergence rate are the same as the gradient descent and the Newton, but are simpler and computationally more efficient than, say the intrinsic Newton and gradient descent algorithms for *spherical weighted averages* in $\mathbf{S}^n$, *cf.* [1].

The rest of this paper is organised as follows. Section 2 describes properties of the Karcher mean. Here we highlight the connection of the Karcher mean with convexity of a set containing the points. The algorithms to compute the Karcher mean in the two spaces, $\mathbb{SO}(n)$ and $\mathbf{S}^n$ are presented in Section 3. Finally, Section 4 concludes the paper.

2 Existence and Uniqueness of the Karcher Mean

It turns out that existence and uniqueness of the Karcher mean depends on the space and its convexity. There are possible problems arising from the curvature of a space. Firstly, since the distance depends on minimizing Riemannian geodesics, one has to make sure that the minimising geodesic exists and is unique — injectivity radius, and then that it lies entirely in the ball containing the set of points Ω — convexity radius. It is also of the interest to know whether the Karcher mean is unique — convexity of Ψ.

Table 1 summarizes geometric properties of the two spaces considered in this paper. The both spaces are symmetric and the unit sphere is a quotient space $\mathbf{S}^n = \mathbb{SO}(n+1)/\mathbb{SO}(n)$. Because the unit sphere is more curved, it poses more theoretical problems related to the existence and uniqueness of the Karcher mean than $\mathbb{SO}(n)$.

Since the paper of Karcher [3], mentioned in the introduction, another significant development has been brought by the work on *convex geometry* by Kendall [4]. There the maximum radius of the convex Riemannian ball containing Ω, for which the Karcher mean exists and is unique, has been improved. For the unit sphere $\mathbf{S}^n$ the radius of the open ball is $\pi/2$. That

[1] We say that $\mathcal{U} \subset \mathbf{M}$ is *convex* if for any $p, q \in \mathcal{U}$ there is a unique in $\mathbf{M}$ minimizing geodesic γ from p to q and $\gamma \subset \mathcal{U}$.

	Special Orthogonal Group $\mathbb{SO}(n)$	**Unit Sphere** $\mathbf{S}^n$
tangent vector	XA — tangent at X, where A — skew-symmetric matrix	V — tangent at x, $\langle V, x\rangle = 0$, product in the ambient $\mathbb{R}^{n+1}$
the inner product	$\langle U, V\rangle := \frac{1}{2}\mathrm{tr}\left(U^{\mathrm{T}}V\right)$, where $U, V \in \mathfrak{so}(n)$	spherical, $\langle U, V\rangle := U^{\mathrm{T}}V$ (in $\mathbb{R}^{n+1}$)
exponential map	$\exp_X(XA) = Xe^A$	$\exp_x V = \cos\|V\|\ x + \frac{\sin\|V\|}{\|V\|}\ V$
log map	$\exp_X^{-1}(P) = X \log X^{\mathrm{T}}P$	$\exp_x^{-1}(p) = (p - x\,\langle x, p\rangle)\ \frac{\arccos\langle x,p\rangle}{\sqrt{1-\langle x,p\rangle^2}}$
curvature	$C(X, Y) = \frac{1}{4}\,\|[X, Y]\|^2$, for $\mathbb{SO}(3)$ C is $\frac{1}{4}$	$C = \frac{1}{R^2}$, for the unit sphere C is 1
injectivity radius[a]	π	π
convexity radius[b]	$\pi/2$	$\pi/2$

[a] for details on how to establish the radius for symmetric spaces *cf.* Kobayashi & Nomizu [5]
[b] follows from the Whitehead's Theorem, *cf.* [5]

Table 1: Comparison of geometric properties of $\mathbb{SO}(n)$ and $\mathbf{S}^n$.

means that if Ω is contained in an open half-sphere, the Karcher mean exists and it is unique. Buss & Fillmore's [1] investigations for the sphere relaxed slightly the conditions allowing the half-sphere to be closed as long as at least one point of Ω lies inside it.

3 Computation of the Karcher Mean

In this section we introduce new iterative methods of calculating Karcher mean on $\mathbb{SO}(n)$ and $\mathbf{S}^n$. They are based on the two standard methods of deriving non-degenerate critical points, namely the gradient descent and the Newton methods. For a smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$ let $x^* \in \mathbb{R}^n$ be a non-degenerate critical point of f, i.e., the Hessian $\mathrm{Hess}\, f(x^*)$ is invertible. Let $\mathfrak{N}_f\colon \mathbb{R}^n \to \mathbb{R}^n$ be given by $x \mapsto \mathfrak{N}_f(x) := x - (\mathrm{Hess}\, f)^{-1}\,\mathrm{grad}\, f(x)$. The *Newton method* is the iteration $x_{i+1} = \mathfrak{N}_f(x_i)$. Similarly, let $\mathfrak{G}_f\colon \mathbb{R}^n \to \mathbb{R}^n$ be given by $x \mapsto \mathfrak{G}_f(x) := x - \mathrm{grad}\, f(x)$. The *gradient descent method* is the iteration $x_{i+1} = \mathfrak{G}_f(x_i)$. The two methods easily extend to Riemannian manifolds. Given a smooth function $f\colon \mathbf{M} \to \mathbb{R}$ and the Riemannian normal coordinates $\varphi_{x_i}\colon \mathbb{R}^n \to \mathbf{M}$ centered at $x_i \in \mathbf{M}$ the iterations

$$x_{i+1} = \varphi_{x_i}\Big(\mathfrak{N}_{f\circ\varphi_{x_i}}(0)\Big) \quad \text{and} \quad x_{i+1} = \varphi_{x_i}\Big(\mathfrak{G}_{f\circ\varphi_{x_i}}(0)\Big) \tag{2}$$

become the intrinsic Newton and gradient methods, respectively. The gradient of the variance of $\Psi\colon \mathbf{M} \to \mathbb{R}$ given by (1) has a particularly simple form $\mathrm{grad}\,\Psi(x) = -\frac{1}{\#\Omega}\sum_{p\in\Omega}\varphi_x^{-1}(p)$, *cf.* [3], and so the intrinsic gradient descent algorithm (2) becomes

$$x_{i+1} = \varphi_{x_i}\left(\frac{1}{\#\Omega}\sum_{p\in\Omega}\varphi_{x_i}^{-1}(p)\right). \tag{3}$$

However, based on the observation in Hüper & Trumpf [2], the same rate of convergence can be obtained with algorithms where Riemannian normal coordinates φ_x are replaced with any,

possibly different, parametrizations agreeing, up to the first derivative, with φ_x. We call these new algorithms the *Newton-like* and the *gradient-like*. We now present these algorithm for the two symmetric spaces, $\mathbb{SO}(n)$ and $\mathbf{S}^n$.

The special orthogonal group

The Riemannian Hessian *operator* $\operatorname{Hess}\Psi\colon \mathcal{T}_X\mathbb{SO}(n) \to \mathcal{T}_X\mathbb{SO}(n)$ of the variance Ψ given by (1) has the following form

$$\operatorname{Hess}\Psi(XA) = \frac{1}{\#\Omega}\sum_{P\in\Omega} \mathcal{R}_{\log(XP^{\mathrm{T}})}\cdot XA, \tag{4}$$

where $\mathcal{R}_X\colon \mathfrak{so}(n) \to \mathfrak{so}(n)$ is defined by

$$Y \mapsto \frac{\operatorname{ad}_X}{2}\coth\frac{\operatorname{ad}_X}{2}\cdot Y \quad \text{with} \quad \mathcal{R}_0(Y) = Y$$

and the linear operator $\operatorname{ad}_X\colon \mathfrak{so}(n) \to \mathfrak{so}(n)$ is given by $\operatorname{ad}_X Y := [X,Y] = XY - YX$.

Theorem 3.1. *Let $\Psi\colon \mathbb{SO}(n) \to I\!R$ be defined by (1).*

1. *The critical points of (1) are precisely the solutions of*

$$\sum_{P\in\Omega}\log(P^{\mathrm{T}}X) = 0. \tag{5}$$

2. *At smooth points of (1), the Riemannian Hessian of (1) is always positive definite.*

Generalized Newton methods Up to our best knowledge there is no known closed form solution of (5) in general. We start this section by proposing Newton-like methods for computing the Karcher mean in $\mathbb{SO}(n)$. Let $X \in \mathbb{SO}(n)$ and consider a smooth parametrization around X $\mu_X\colon \mathbb{R}^{n(n-1)/2} \to \mathbb{SO}(n)$ with $\mu_X(0) = X$, i.e., a local diffeomorphism around $0 \in \mathbb{R}^{n(n-1)/2}$. If there exists an open neighborhood $\mathcal{U} \subset \mathbb{SO}(n)$ of X^* and a smooth map

$$\mu\colon \mathcal{U}\times\mathbb{R}^{n(n-1)/2} \to \mathbb{SO}(n)$$

such that $\mu(X,v) = \mu_X(v)$, for all $X \in \mathcal{U}$ and $v \in \mathbb{R}^{n(n-1)/2}$, we will call $\{\mu_X\}_{X\in\mathbb{SO}(n)}$ a *locally smooth family of parametrizations around X^**. We now present three different families of this kind. In what follows we will identify $\mathfrak{so}(n) \cong \mathbb{R}^{n(n-1)/2}$.

$$\mu^{\mathrm{GS}}\colon \mathcal{U}\times\mathfrak{so}(n) \to \mathbb{SO}(n) \qquad \text{given by} \qquad (X,A) \mapsto Q\left(X(\mathbf{I}+A)\right), \tag{6}$$

$$\mu^{\mathrm{cay}}\colon \mathcal{U}\times\mathfrak{so}(n) \to \mathbb{SO}(n) \qquad \text{given by} \qquad (X,A) \mapsto X\left(\mathbf{I}+\frac{A}{2}\right)\left(\mathbf{I}-\frac{A}{2}\right)^{-1}, \tag{7}$$

$$\mu^{\exp}\colon \mathcal{U}\times\mathfrak{so}(n) \to \mathbb{SO}(n) \qquad \text{given by} \qquad (X,A) \mapsto X\exp A, \tag{8}$$

here $Q(X(\mathbf{I}+A))$ denotes the Q-factor of the unique QR-decomposition of the matrix $X(\mathbf{I}+A)$, i.e., the orthogonal matrix produced from the Gram-Schmidt process applied to the columns of $X(\mathbf{I}+A)$. The following Lemma is easily verified.

Lemma 3.1. *Let $\mathcal{U} \subset \mathbb{SO}(n)$ be an open neighborhood of $X \in \mathbb{SO}(n)$. Let $\mu_X^{\mathrm{GS}}, \mu_X^{\mathrm{cay}}$, and $\mu_X^{\exp}$ be defined as (6), (7), and (8), respectively. Then for arbitrary $A \in \mathfrak{so}(n) \cong \mathbb{R}^{n(n-1)/2}$*

$$\mathrm{D}\,\mu_X^{\exp}(0) \cdot A = \mathrm{D}\,\mu_X^{\mathrm{cay}}(0) \cdot A = \mathrm{D}\,\mu_X^{\mathrm{GS}}(0) \cdot A = XA. \tag{9}$$

Algorithm 3.1 (Newton-like Method)**.** *Given a final set of points $\Omega \subset \mathbb{SO}(n)$, compute a local minimum of Ψ given by (1).*

Step 1 *Set $X \in \mathbb{SO}(n)$ to an initial estimate of the Karcher mean, such as to any point of Ω.*

Step 2 *Compute* $\dfrac{1}{\#\Omega} \displaystyle\sum_{P \in \Omega} \log(P^{\mathrm{T}} X)$.

Step 3 *Stop if* $\left\| \dfrac{1}{\#\Omega} \displaystyle\sum_{P \in \Omega} \log(P^{\mathrm{T}} X) \right\|$ *is sufficiently small.*

Step 4 *Solve the linear equation* $\operatorname{Hess} \Psi(XA) = -\operatorname{grad} \Psi(X)$ *for A.*

Step 5 *Set* $X := \mu_X(A)$.

Step 6 *Goto* **Step 2**.

We have the following result.

Theorem 3.2. *If Algorithm 3.1 converges then it converges locally quadratically fast.*

The unit sphere

From now on we assume that Ω is contained in an open half-sphere. For any $x \in \mathbf{S}^n$ and $V \in \mathcal{T}_x \mathbf{S}^n$ the Riemannian Hessian of the variance $\Psi \colon \mathbf{S}^n \to \mathbb{R}$ is given by

$$\operatorname{Hess} \Psi(V, V) = \frac{1}{\#\Omega} \sum_{p \in \Omega} \left(\frac{1}{\sin^2 \theta_p} \left(1 - \theta_p \cot \theta_p\right) \langle V, p \rangle^2 + \theta_p \cot \theta_p \, \|V\|^2 \right),$$

where $\theta_p = \operatorname{dist}(x, p) = \arccos \langle x, p \rangle$. There are the two important properties of $\operatorname{Hess} \Psi$ that are essential to analysis of convergence of the algorithms (2).

Lemma 3.2. *The Hessian of Ψ satisfies* $\operatorname{Hess} \Psi(V, V) \leq \|V\|^2$, *for any $x \in \mathbf{S}^n$ and $V \in \mathcal{T}_x \mathbf{S}^n$.*

Lemma 3.3. *The Hessian $\operatorname{Hess} \Psi$ is positive definite at the Karcher mean $\bar{x}$. Therefore there exists a neighbourhood $\mathcal{U} \subset \mathbf{S}^n$ of $\bar{x}$, where Ψ is convex.*

The intrinsic gradient descent and the Newton algorithms in the sphere studied in Buss & Fillmore [1] are given by (2). These algorithms are computationally expensive because at each iteration they require calculations of normal coordinates of every point of Ω, *cf.* (3). As in the case of $\mathbb{SO}(n)$, we may replace the normal coordinates φ_x with another mapping, and preserve the rate of convergence. For the sphere a good choice is the orthogonal projection

$$\pi_x \colon \mathbb{R}^{n+1} \to \mathcal{T}_x \mathbf{S}^n \quad \text{given by} \quad z \mapsto \pi_x(z) := z - \langle x, z \rangle \, x = (\mathbf{I} - xx^{\mathrm{T}}) z.$$

We conclude this section with a new gradient-like algorithm.

Generalized Gradient Method The gradient-like algorithm (10) in the unit sphere $\mathbf{S}^n$ converges to the Karcher mean locally linearly fast

$$x_{i+1} = \varphi_{x_i}\left(\frac{1}{\#\Omega}\sum_{p\in\Omega}\pi_{x_i}(p)\right). \tag{10}$$

4 Conclusion

This paper presents, to the best of our knowledge, the first efficient and reliable numerical algorithm for computing the Karcher mean of points in $\mathbb{SO}(n)$. Specifically, under the same necessary condition which ensures the Karcher mean is well defined, the algorithm is proved to converge to the Karcher mean. Moreover, the local rate of convergence is quadratic. Finally, a new gradient-like algorithm in $\mathbf{S}^n$ is presented.

References

[1] Samuel R. Buss and Jay P. Fillmore. Spherical averages and applications to spherical splines and interpolation. *ACM Transactions on Graphics*, 20(2):95–126, 2001.

[2] Knut Hüper and Jochen Trumpf. Newton-like methods for numerical optimization on manifolds. In *Proceedings of the 38 Asilomar Conference on Signals, Systems and Computers, November 7–10 2004, California, USA*, volume 1, pages 136–139, 2004.

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

[4] Wilfrid S. Kendall. Probability, convexity and harmonic maps with small image I: uniqueness and fine existence. *Proceedings of the London Mathematical Society*, 61:371–406, 1990.

[5] Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of Differential Geometry*, volume 2 of *Interscience tracts in pure and applied mathematics*. Interscience Publishers, New York, 1969.

[6] Maher Moakher. Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):1–16, 2002.