ALGORITHM COMPARISON FOR KARCHER MEAN COMPUTATION OF ROTATION MATRICES AND DIFFUSION TENSORS

Quentin Rentmeesters, P.-A. Absil

Université catholique de Louvain,
Department of Mathematical Engineering,
B-1348 Louvain-la-Neuve, Belgium

ABSTRACT

This paper concerns the computation, by means of gradient and Newton methods, of the Karcher mean of a finite collection of points, both on the manifold of 3×3 rotation matrices endowed with its usual bi-invariant metric and on the manifold of 3×3 symmetric positive definite matrices endowed with its usual affine invariant metric. An explicit expression for the Hessian of the Riemannian squared distance function of these manifolds is given. From this, a condition on the step size of a constant step gradient method that depends on the data distribution is derived. These explicit expressions make a more efficient implementation of the Newton method possible and it is shown that the Newton method outperforms the gradient method in some cases.

1. INTRODUCTION

In many signal processing applications, it is necessary to deal with data that belong to a Riemannian manifold. For instance, in computational anatomy [1, 2], the data are diffusion tensors or symmetric positive definite matrices representing the diffusion of water at some voxel in the white matter of the brain. In computer vision and robotics, the orientation of an object is often represented by a rotation with respect to a reference frame, see [3, 4]. To reduce a measurement noise on these data or simply to obtain a measure of centrality, an efficient technique to compute the mean of a set of points is required.

A notion of mean on Riemannian manifolds has been introduced by Karcher in [5]. A unit step size gradient descent method to compute this mean has been proposed in [1] for the set of symmetric positive definite matrices. In [6], a proof of convergence is given under some conditions on the curvature of the manifold but no step size selection rule is proposed to ensure the convergence. On Lie groups, such as the set of rotation matrices, a unit step size gradient method with a convergence proof is proposed in [7] and a Newton method is introduced in [8].

In this paper, the potential interest of using a Newton method to compute the Karcher mean is discussed. Such a technique has an asymptotic quadratic rate of convergence. So in practice, a Newton method requires less iterations than first order techniques to reach a sufficiently high accuracy. But the cost per iteration is significantly higher since the Hessian of the objective function must be computed and the symmetric linear system of Newton's equations must be solved at each iteration. Consequently, a gain in terms of computational time is not guaranteed. Here, explicit expressions for the eigenpairs of the curvature endomorphism are derived and, from this, an efficient implementation of the Newton method presented in [8] is proposed. Using this efficient implementation, it is shown that the Newton method is faster than the gradient technique when a high accuracy is required, if the data points are not too concentrated around one point. A convergence analysis for the gradient method on the set of symmetric positive definite matrices is also given. More precisely, a condition on the step size, related to the data distribution, is introduced to ensure the convergence. Our analysis can be seen as a particularization of the result of [7], for the set of rotations but, to the best of our knowledge, this analysis is new for the set of symmetric positive definite matrices.

The paper is organized as follows. Section 2 introduces the Riemannian geometry of the set of rotations in the 3-dimensional space $SO(3)$ and the set of diffusion tensors of dimension 3 denoted by P_3^+. Some results about the uniqueness of the Karcher mean are recalled in section 3. An explicit expression for the Hessian of the Riemannian squared distance function is given in section 4 and from this a condition on the step size of the gradient technique is deduced to ensure the convergence, see section 5. Section 6 gives some details about the implementation and section 7 compares the gradient and the Newton technique in terms of computational time.

2. RIEMANNIAN GEOMETRY OF $SO(3)$ AND P_3^+

In this section, the Riemannian geometries of $SO(3)$ and P_3^+ are briefly described. An element of $SO(3)$ can be represented by an orthogonal matrix X of determinant 1:

$$SO(3) = \{X \in \Re^{3\times 3} | X^\top X = I,\ \det(X) = 1\}.$$

This representation is particularly useful since the application of the rotation X to a vector $v \in \Re^3$ is given by Xv. Notice that it is also possible to use unit norm quaternions or an angle and an axis to represent a rotation. $SO(3)$ has a Lie group structure and its Lie algebra $so(3)$, or the tangent space at the identity, is given by the set of skew-symmetric matrices

$$so(3) = \{X \in \Re^{3\times 3} | X^\top = -X\}.$$

Notice that $so(3)$ is also equivalent to $\Re^3$ endowed with the cross product, denoted by $\times$, for the Lie bracket. Table 1 summarizes the two representations. By introducing the following inner product on $so(3)$

$$g(X,Y) = \frac{1}{2}\text{trace}(X^\top Y),$$

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its author(s).

Vector	Matrix
$\tilde{v} = (v_1, v_2, v_3) \in \Re^3$	$v = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}$
$g_p(\tilde{v}, \tilde{v}) = \|\tilde{v}\|_2$	$g_p(v, v) = \frac{1}{2}\text{trace}(v^\top v)$
$[\tilde{v}, \tilde{w}] = \tilde{v} \times \tilde{w}$	$[v, w] = vw - wv$

Table 1: Representation of $so(3)$

we turn $SO(3)$ into a Riemannian manifold. This inner product is bi-invariant (on the left and on the right) to the action of the orthogonal group and $SO(3)$ has the structure of a Riemannian symmetric space [9].

Let P_3^+ be the set of symmetric positive definite matrices

$$P_3^+ = \{P \in \Re^{3\times 3} | P = P^\top, P \succ 0\}.$$

This is also the set of ellipsoids in $\Re^3$ and it is also called the set of diffusion tensors [2]. This set is a homogeneous space: the general linear group $GL(3) = \{g \in \Re^{3\times 3} | \det(g) \neq 0\}$ acts on it transitively by congruence $P \mapsto gPg^\top$. The tangent space at p is the set of symmetric matrices:

$$T_p P_n^+ = \{V \in \Re^{n\times n} | V = V^\top\}.$$

When this space is endowed with the Riemannian metric

$$g_p(X, Y) = \text{trace}(Xp^{-1}Yp^{-1}), \tag{1}$$

called the affine invariant metric, P_3^+ becomes a Riemannian symmetric space [2].

Using these metrics, we turn these manifolds $\mathscr{M}$ into metric spaces by introducing the following distance function

$$d(x, y) = \underset{\gamma(t)\in\mathscr{M} \text{ s.t. } \gamma(0)=x,\ \gamma(1)=y}{\text{argmin}} \int_0^1 \sqrt{g(\dot{\gamma}(s), \dot{\gamma}(s))} ds.$$

This distance is the length of the minimizing geodesic $t \mapsto \gamma(t)$ starting at x and ending at y. Since $SO(3)$ is a Lie group under matrix multiplication, for all $A, B \in SO(3)$, there exists a $C \in SO(3)$ such that $A = CB$. The Riemannian distance corresponds to the angle of the rotation represented by the matrix C. Notice that the Riemannian distance cannot exceed π since a rotation whose angle of rotation is $\pi + \theta$, $0 < \theta \leq \pi$ is equivalent to a rotation of angle $\theta - \pi$ around the same axis.

On P_3^+, the Riemannian distance between the identity tensor and another tensor P is

$$d(I, P) = \sqrt{\log(d_1)^2 + \log(d_2)^2 + \log(d_3)^2}$$

where $P = U\text{diag}(d_1, d_2, d_3)U^\top$ (eigenvalue decomposition). Consequently, an ellipsoid on the border of the cone, i.e. a rank deficient ellipsoid, is at an infinite distance to the identity. Notice that endowed with this distance function, $SO(3)$ and P_3^+ are complete metric spaces.

3. KARCHER MEAN

Let $q_1, ..., q_n$ be a set of n points on the manifold $\mathscr{M}$ and let

$$F : \mathscr{M} \to R_+; \qquad x \mapsto \frac{1}{2n} \sum_{i=1}^n d(x, q_i)^2. \tag{2}$$

A point $x \in \mathscr{M}$ is a Karcher mean, see [5], if $\text{grad}F(x) = 0$, i.e. x is a stationnary point of F. Notice that in the literature, other denominations are used: centroid, barycenter, Riemannian center of mass.

In general, the Karcher mean is not unique. But on the Riemannian manifolds with positive sectional curvature κ and injectivity radius $\text{inj}(\mathscr{M})$, the Karcher mean is unique in the open ball $B(x_0, \frac{1}{2}\min\{\text{inj}(\mathscr{M}), \frac{\pi}{\sqrt{\kappa}}\})$ if all the data points belong to that ball. Since the sectional curvature of $SO(3)$ is $1/4$ and its injectivity radius $\text{inj}(SO(3)) = \pi$, the radius of the ball is $\pi/2$. On P_3^+, the sectional curvature is negative and the Karcher mean is unique if no data point lies on the boundary of the cone. These results are due to Kendall, see [10].

To find the Karcher mean inside such a ball, a gradient technique can be used. The gradient of the function (2), see [5], is

$$\text{grad}F(\mu) = \frac{1}{n} \sum_{i=1}^n \exp_\mu^{-1}(q_i), \tag{3}$$

where $\exp_x^{-1}(y)$ stands for the log-mapping that returns a tangent vector v at x such that the geodesic curve $t \mapsto \gamma(t)$ with $\gamma(0) = x$ and $\dot{\gamma}(0) = v$ satisfies $\gamma(1) = y$. In [6], a gradient method is proved to converge under some assumptions on the curvature of the manifold. This is, to the best of our knowledge, the most recent contribution that proves the convergence of the method in a general setting. The algorithm is the following:

Algorithm 1 Gradient technique

1: Given a set of n points $q_i \in \mathscr{M}$ for $1 \leq i \leq n$ and an initial guess for the Karcher mean μ_0;
2: Set $\mu_k = \mu_0$ and $k = 1$;
3: **until** $g_{\mu_k}(\text{grad}F(\mu_k), \text{grad}F(\mu_k)) < \varepsilon$ **do**
4: Compute the gradient of F at μ_k using (3);
5: Move along the geodesic curve starting at μ_k: $\mu_{k+1} = \exp_{\mu_k}(-h_{\mu_k}\text{grad}F(\mu_k))$;
6: Set $k = k + 1$;
7: **end**
8: **return** μ_k (the estimation of the Karcher mean)

In this algorithm the step size h_{μ_k} has to be chosen using for instance a line search technique. But a line search technique requires to evaluate the objective function, which is as expensive as computing the gradient. One possibility is to fix the step size in advance. For instance we can fix $h_{\mu_k} = 1$. This strategy is followed in [7] for $SO(3)$ where a global convergence result is given if the data points are located inside an open ball of radius $\pi/2$. Notice that this result is valid on any compact Lie group. The same strategy is used in [1] to compute the Karcher mean on P_3^+. But to the best of our knowledge, there is no convergence proof for P_3^+.

4. EXPLICIT EXPRESSION FOR THE HESSIAN

On P_3^+ and $SO(3)$, it is possible to have an explicit formula for the Hessian of the Riemannian squared distance function and to derive bounds on the eigenvalues of the Hessian in function of the data distribution.

Let p and q be two points in $\mathscr{M}$, $\Delta p \in T_p\mathscr{M}$ and $s \to c(s)$, a smooth curve in $\mathscr{M}$ such that $c(0) = p$ and $\dot{c}(0) = \Delta p$; the

Hessian of the Riemannian squared distance function

$$f_q(p) = \frac{1}{2} d(p,q)^2$$

is given by

$$\begin{aligned} \mathrm{Hess} f_q(p)[\Delta p, \Delta p] &\triangleq \frac{d}{ds^2} f_q(c(s))|_{s=0}, && (4) \\ &= g_p(\dot{J}(1), \Delta p), && (5) \end{aligned}$$

where $\dot{J}(1) = D_t J(t)|_{t=1}$ is the covariant derivative of the Jacobi field $J(t)$ along the geodesic curve $t \mapsto \gamma(t)$ joining q and p such that $\gamma(0) = q$ and $\gamma(1) = p$, see [5], that satisfies

$$\begin{aligned} D_t^2 J(t) + R(J(t), \dot{\gamma}(t))\dot{\gamma}(t) = 0, \quad (6) \\ J(0) = 0 \in T_q\mathscr{M}, \\ J(1) = \Delta p \in T_p\mathscr{M}. \end{aligned}$$

Since the curvature endomorphism R is parallel on symmetric spaces, this implies that (6) is a constant coefficient differential equation and it can be solved explicitly, see [11, 9, 8] for more details. This yields,

$$\mathrm{Hess} f_q(p)[\Delta p, \Delta p] = g_p(\underbrace{\sum_{k=1}^{m} g_p(\Delta p, E_k(1)) w_k(1) E_k(1)}_{\dot{J}(1)}, \Delta p), \quad (7)$$

$$= \sum_{k=1}^{m} g_p(\Delta p, E_k(1))^2 w_k(1), \quad (8)$$

with

$$w_k(t) = \begin{cases} \sqrt{\lambda_k}\cot(t\sqrt{\lambda_k}) & \text{if } \lambda_k > 0, \\ 1/t & \text{if } \lambda_k = 0, \\ \sqrt{-\lambda_k}\coth(t\sqrt{-\lambda_k}) & \text{if } \lambda_k < 0, \end{cases}$$

where the λ_k's are the eigenvalues of the Jacobi operator

$$J \mapsto R(J, \dot{\gamma}(0))\dot{\gamma}(0), \quad (9)$$

and the E_k's are its corresponding eigenvectors. Since the Jacobi operator (9) is a symmetric linear map, it admits an orthonormal basis of eigenvectors. Let us assume that $E_1, ..., E_m$ form an orthonormal basis at $q = \gamma(0)$ that can be extended to a frame $E_1(t), ..., E_m(t)$ by parallel transport along the curve $t \mapsto \gamma(t)$. Notice that $\dot{\gamma}(0) = \exp_q^{-1}(p)$ and $\|\dot{\gamma}(0)\| = d(p,q)$. To compute the eigenpairs of the Jacobi operator (9), a representation of the tangent vectors $\dot{\gamma}(0)$ and J is required. The representation of $\dot{\gamma}(0)$ is denoted by A and the representation of J by X.

On $SO(3)$, an element of the Lie algebra can be represented by a 3×3 skew-symmetric matrix. Using this representation, the curvature endomorphism is given by (see [12])

$$R(X,Y)Z = \frac{1}{4}[Z, [X,Y]], \quad (10)$$

where $X, Y, Z \in so(3)$ are skew symmetric matrices. But, using the vector representation of the Lie algebra, see Table 1 we can write

$$R(X,A)A = \frac{1}{4} A \times (X \times A),$$

	$SO(3)$	P_3^+
$\lambda_{\min}$	0	$-\frac{1}{4}(\sigma_{\max} - \sigma_{\min})^2$
$\lambda_{\max}$	$d(p,q)^2/4$	0
$\min_k w_k(1)$	$\frac{d(p,q)}{2}\cot(\frac{d(p,q)}{2})$	1
$\max_k w_k(1)$	1	$\frac{\sigma_{\max}-\sigma_{\min}}{2}\coth(\frac{\sigma_{\max}-\sigma_{\min}}{2})$

Table 2: Largest $\lambda_{\max}$ and smallest $\lambda_{\min}$ eigenvalues of the Jacobi operator (9) and bounds on the Hessian of f_q at p. On $SO(3)$, $d(p,q) = \|\exp_q^{-1}(p)\|$ is the Riemannian distance between p and q. On P_3^+, $\sigma_{\max}$ and $\sigma_{\min}$ are the largest and smallest eigenvalues of A, the representation of $\dot{\gamma}(0) = \exp_q^{-1}(p)$.

where $A \in \Re^3$ represents $\dot{\gamma}(0)$ and $X \in \Re^3$ represents the Jacobi field J. Consequently, the eigenpairs of the Jacobi operator (9) with the vector representation of the Lie algebra are described by

- $\lambda = 0$, E_1;
- $\lambda = \frac{\|A\|^2}{4}$, E_2, E_3;

where $E_1 = \frac{A}{\|A\|}$ and $[E_1, E_2, E_3]$ is a positively oriented orthonormal basis of $\Re^3$.

On P_3^+, the curvature endomorphism at the identity is also given by (10) where X, Y, Z are tangent vectors at the identity, i.e. symmetric matrices, see [8] for more details. Since we can 'translate' the problem to the identity using an isometry, only the diagonalization of the curvature endomorphism at the identity is required and we can assume that $q = I$. Once again let A be a representation of $\dot{\gamma}(0)$. Since A is a symmetric matrix, it can be diagonalized by an orthogonal transformation, i.e. $A = U\Sigma U^\top$ where $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$. Let e_i be the i-th identity vector. The eigenvalues λ and related eigenvectors E of (9) are

- $\lambda = 0$, $E = U e_i e_i^\top U^\top$ with $1 \le i \le 3$ (3 eigenpairs of this type);
- $\lambda = -\frac{1}{4}(\sigma_i - \sigma_j)^2$, $E = U\frac{1}{\sqrt{2}}(e_i e_j^\top + e_j e_i^\top)U^\top$ with $1 \le i < j \le 3$ (3 eigenpairs of this type).

The proof is omitted and will appear elsewhere. From these explicit expressions of the eigenpairs of the curvature endomorphism, it is possible to derive bounds on the Hessian of f and since $F(x) = \frac{1}{n}\sum_{i=1}^n f_{q_i}(x)$ with $f_{q_i}(x) = \frac{1}{2}d(x,q_i)^2$, these bounds are also valid for F. From (8), we have:

$$\min_k w_k(1)\|\Delta p\|^2 \le \mathrm{Hess} f_q(p)[\Delta p, \Delta p] \le \max_k w_k(1)\|\Delta p\|^2. \quad (11)$$

The term $w_k(1)$ is shown in Figure 1. On P_3^+, the sectional curvature is negative, i.e. $\lambda \le 0$, so $w_k(1)$ is greater or equal to 1 and the Hessian is positive definite. On $SO(3)$, the sectional curvature is positive, i.e. $\lambda \ge 0$, and $w_k(1)$ is upper bounded by 1 and is positive if $\sqrt{\lambda} < \pi/2$. Table 2 gives upper and lower bounds on the eigenvalue of the Hessian of f_q at p in function of the Riemannian distance $d(p,q)$ on $SO(3)$ and in function of the largest $\sigma_{\max}$ and the smallest $\sigma_{\min}$ eigenvalues of A on P_3^+.

Observe that this development confirms that the distance function is convex if $d < \pi$ on $SO(3)$. So, the Karcher mean is unique on a ball of radius $\pi/2$ as shown in [7]. And if the

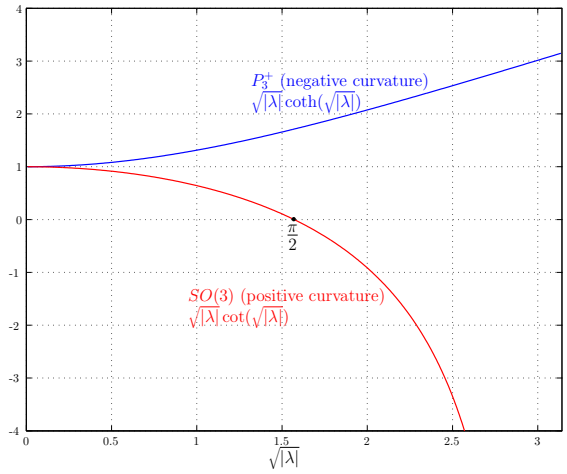

Figure 1: Graphic of $w_k(1)$ in function of $\sqrt{|\lambda|}$ in the positive curvature case (red) and in the negative curvature case (blue)

data belongs to a ball of radius $\pi/2 \leq r \leq \pi$ centered at p, then the Karcher mean is still unique inside the ball of radius $\pi - r$. In fact, by the triangle inequality, the distance between any point inside this ball and the data points cannot exceed π, see [3].

5. CHOICE OF THE STEP SIZE

This section proposes a convergence analysis of the gradient technique (Algorithm 1) with constant step size. This follows the ideas of [13] and [6].

The proof of convergence of the constant step algorithm 1 is based on the relaxation principle, i.e. the value of the cost function at each iteration must be strictly decreasing. So, the step size h must be chosen to ensure this decrease. Let μ_k be the current iterate and let $h \mapsto \gamma(h) = \mu_{k+1}$ be the geodesic curve such that $\gamma(0) = \mu_k$ and $\dot{\gamma}(0) = -\text{grad}F(\mu_k)$. Using a Taylor expansion we obtain

$$F(\gamma(h)) = F(\gamma(0)) - h g_{\gamma(0)}(\text{grad}F(\gamma(0)), \text{grad}F(\gamma(0))) + \frac{h^2}{2}\text{Hess}F(\gamma(h^*))[\dot{\gamma}(h^*), \dot{\gamma}(h^*)],$$

with $h^* \in]0,h[$. Let $L > 0$ such that $\text{Hess}F(\gamma(h^*))[\dot{\gamma}(h^*), \dot{\gamma}(h^*)] \leq L\|\dot{\gamma}(h^*)\|^2 = L\|\dot{\gamma}(0)\|^2 = L\|\text{grad}F(\gamma(0))\|^2$, we have

$$F(\gamma(h)) - F(\gamma(0)) \leq (-h + \frac{1}{2}h^2 L)\|\text{grad}F(\gamma(0))\|^2.$$

To guarantee the decrease of the objective function, the upper bound on the right hand side must be strictly negative. This is true if $0 < h < 2/L$, and $h = 1/L$ minimizes the bound. By taking the sum from 1 to k on the left-hand side and on the right-hand side, one obtains

$$\sum_{i=1}^{k} \|\text{grad}F(\mu_i)\|^2 \leq \frac{(F(\mu_{k+1}) - F(\mu_1))}{(-h + \frac{1}{2}h^2 L)},$$

which implies $\|\text{grad}F(\mu_i)\| \to 0$ when $k \to \infty$ since F is bounded from below.

Notice that our upper and lower bounds on the Hessian, see Table 2, depend on the distribution of the data points. Thus we have to assume that all data points lie (strictly) inside a ball, to have an estimation of L and so to choose a step size h that ensures the convergence. Consequently, the step size must be chosen to ensure that we do not escape from this ball. This is guaranteed if $h \in]0, 1/L]$. In fact, let $\bar{h}$ be such that $\gamma(\bar{h})$ is on the border of the ball. Since all the data points lie strictly inside the ball, $\text{grad}F(\gamma(\bar{h}))$ points outside the ball and $F(\gamma(h))$ is increasing at this point. Since we assume that F is convex inside the ball, there is one and only one point $h' \in]0, \bar{h}[$ with $\gamma(h')$ inside the ball such that $\frac{d}{dh}F(\gamma(h))|_{h=h'} = 0$. If $h \leq 1/L$ then $\frac{d}{dh}F(\gamma(h))|_h < 0$ and so $h < h'$ which ensures that $\gamma(h)$ is inside the ball.

On $SO(3)$, $L = 1$; see (11) and Table 2. On P_3^+, L depends on the distribution of the data points. To get more insight on the value of L, it is easier to relate it to the condition number of the ellipsoids. Given $P \in P_3^+$, one can write $P = UDU^\top$ (eigenvalues decomposition), and let $c = \frac{d_{\max}}{d_{\min}}$ be its condition number with respect to the 2-norm; then $L = \max_k w_k(1)$ is given by

$$L(c) = \frac{1}{2}\log(c)\frac{c+1}{c-1}. \tag{12}$$

This function tends to 1 when c is going to 1 and is strictly increasing for $c > 1$. So if the condition number of the ellipsoids are close to 1, then the upper bound L will be a bit larger than 1 and the Hessian will be close to the identity.

6. IMPLEMENTATION

Given $p \in SO(3)$, a tangent vector at p can be represented by a matrix of the form pv where v is skew-symmetric, see the definition of the tangent space. But it is also possible to represent only $v \in T_I SO(3)$ without forgetting that this tangent vector at the identity or in the Lie algebra is related to the point p. This approach is computationally more efficient since it does not require to compute and store the matrix pv. Furthermore, to represent an element of the Lie algebra, a vector $\tilde{v} \in \Re^3$ is used, see Table 1. Using this representation and the Rodrigues formulas, we have

- $\exp_p(v) : so(3) \to SO(3)$,
  $\exp_p(v) = p\text{Exp}(v) = p(I + \frac{\sin\|\tilde{v}\|}{\|\tilde{v}\|}v + \frac{(1-\cos(\|\tilde{v}\|)}{\|\tilde{v}\|^2}v^2)$,
  and if $\|\tilde{v}\| < 10^{-15}$, set $\exp_p(v) = p$
- $\exp_p^{-1}(q) : SO(3) \to so(3)$,
  $\exp_p^{-1}(q) = \text{Log}(p^\top q) = \frac{\theta}{2\sin\theta}\begin{bmatrix} -R(1,2)+R(2,1) \\ R(1,3)-R(3,1) \\ -R(2,3)+R(3,2) \end{bmatrix}$,
  where $R = p^\top q$ and $\theta = \arccos\frac{\text{trace}(R)-1}{2}$.
  and if $|\theta| < 10^{-15}$, set $\exp_p^{-1}(q) = 0_3$

where Exp and Log stand for the matrix exponential and the matrix logarithm.

On P_3^+ the exponential and logarithm mappings are given by, see [2],

$$\exp_p(v) = p^{1/2}\text{Exp}(p^{-1/2}vp^{-1/2})p^{1/2},$$
$$\exp_p^{-1}(q) = p^{1/2}\text{Log}(p^{-1/2}qp^{-1/2})p^{1/2}.$$

r vs n	4	10	100	1000
$\pi/4$	(4,10)	(5,11)	(3,11)	(5,10)
$\pi/2$	(24,12)	(29,7)	(29,4)	(28,5)
$3\pi/4$	(34,14)	(45,7)	(46,3)	(43,3)

Table 3: Mean (first entry) and standard deviation (second entry) of the time reduction in % in function of the number of data points n and the radius of the ball r on $SO(3)$

Notice that the action $x \mapsto p^{-1/2}xp^{-1/2}$ maps p to the identity I and its differential is the map $X_x \mapsto p^{-1/2}X_xp^{-1/2}$ where X_x is a tangent vector at x. This action is also an isometry with respect to (1). In the computation, at each iteration, the current estimate of the mean μ_k in Algorithm 1 is mapped to the identity using the isometry $x \mapsto {\mu_k}^{-1/2}x{\mu_k}^{-1/2}$ and the same isometry is applied to each data point. This isometry is stored in memory. So, at the end, the Karcher mean is computed by applying the inverse of the composition of all these isometries.

With our explicit expressions for the eigenpairs of the curvature endomorphism, one can express the Hessian form (8). This leads to an efficient implementation of the Newton method that can outperform the gradient method in some cases. To implement the Newton method, a basis is needed. On $SO(3)$ one can choose the canonical basis of $\Re^3$ and on P_3^+ the basis $e_ie_i^\top$ $1 \leq i \leq 3$; $\frac{1}{\sqrt{2}}(e_ie_j^\top + e_je_i^\top)$ $1 \leq i < j \leq 3$. Then using (8), one obtains a representation of the Hessian matrix in the chosen basis. Notice also that, since the current estimation of the mean is mapped back to I at each iteration, the diagonalization of the curvature endomorphism is only required at the identity. The details of the implementation will appear elsewhere.

7. NUMERICAL RESULTS

A set of n points was generated around the identity, w.l.o.g. since $SO(3)$ and P_3^+ are homogeneous spaces. The points are distributed uniformly (radially and in each direction) on a ball of radius r. The time required to execute the gradient and Newton algorithms to reach a given accuracy was computed using the *tic; ... toc;* function of Matlab and averaged over 100 runs. To be independent of the computer power, we only report the time reduction $(1 - \frac{T_{\text{Newton}}}{T_{\text{gradient}}})100$ (in %). This set of points was generated 100 times for different values of the radius r. The mean and the standard deviation of the time reduction is presented in Table 3 for $SO(3)$ and in Table 4 for P_3^+.

On $SO(3)$, the accuracy was set to $\|\text{grad}F(\mu_k)\| < 10^{-15}$. The unit step size gradient technique was used since it guarantees the convergence. Notice that in the case $r = 3\pi/4$, the initial condition was picked randomly inside the ball of radius $\pi/4$ since the Karcher mean is unique inside of this ball, see Section 4. For $r = \pi/4$, the Newton method does not perform better than the gradient method but for $r = \pi/2$ and $r = 3\pi/4$, one observes a significant improvement with the Newton method.

On P_3^+, the accuracy is set to 10^{-10}. The Newton method was compared to a unit time step gradient method. This step size does not guarantee the convergence but it is used in [1] and [2]. The Newton method performs better if the radius of the ball is larger than 3. For $r \geq 4$, the gradient technique does not converge all the time and it is stopped if the number of gradient iterations exceeds 200. To avoid this, it is possible to reduce the step size in order to ensure the convergence, see Section 5. But most of the time, even with the best step size, i.e. the one that minimizes the number of iterations, the Newton method is faster.

r vs n	4	10	100	1000
1	(1,15)	(-13,12)	(-30,11)	(-33,10)
2	(21,16)	(5,10)	(-17,11)	(-21,14)
3	(38,18)	(27,13)	(10,6)	(6,7)
4	(58,14)	(43,11)	(27,8)	(25,5)
5	(70,13)	(56,12)	(42,7)	(36,6)

Table 4: Mean (first entry) and standard deviation (second entry) of the time reduction (in %) in function of the number of data points n and the radius of the ball r on P_3^+

8. CONCLUSIONS

Explicit expressions for the eigenpairs of the curvature endomorphism on $SO(3)$ and P_3^+ were given. Based on this, an efficient Newton method for computing the Karcher mean of a finite set of points was proposed and compared to the gradient technique studied in [7] and [6]. A condition on the data distribution was also given to ensure the convergence of this gradient method. The same analysis can be carried out on other symmetric spaces like $SO(n)$ and P_n^+ for larger n or the Grassmann manifold.

REFERENCES

[1] X. Pennec, "Statistical computing on manifolds: from riemannian geometry to computational anatomy," *Emerging Trends in Visual Computing*, pp. 347–386, 2009.

[2] P.T. Fletcher and S. Joshi, "Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors," *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, pp. 87–98, 2004.

[3] R. Hartley, J. Trumpf, and Y. Dai, "Rotation averaging and weak convexity," .

[4] J. Lee and S.Y. Shin, "General construction of time-domain filters for orientation data," *IEEE Transactions on Visualization and Computer Graphics*, pp. 119–128, 2002.

[5] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on pure and applied mathematics*, vol. 30, no. 5, pp. 509–541, 1977.

[6] H. Le, "Estimation of Riemannian barycentres," *LMS Journal of Computation and Mathematics*, vol. 7, pp. 193–200, 2004.

[7] J.H. Manton, "A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups," in *Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th*. IEEE, 2005, vol. 3, pp. 2211–2216.

[8] R. Ferreira, J. Xavier, J. Costeira, and V. Barroso, "Newton algorithms for Riemannian distance related problems on connected locally symmetric manifolds," Tech. Rep., Technical Report, Signal and Image Processing Group (SPIG), Institute for Systems and Robotics (ISR), 2008.

[9] J. Cheeger and D.G. Ebin, *Comparison theorems in Riemannian geometry*, American Mathematical Society, 2008.

[10] W.S. Kendall, "Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence," *Proceedings of the London Mathematical Society*, vol. 3, no. 2, pp. 371, 1990.

[11] M.P. Do Carmo, *Riemannian geometry*, Birkhauser, 1992.

[12] B. O'neill, *Semi-Riemannian geometry: with applications to relativity*, Academic Pr, 1983.

[13] Y. Nesterov, "Introductory lectures on convex programming," *Lecture Notes*, pp. 119–120, 1998.