NEWTON METHOD FOR RIEMANNIAN CENTROID COMPUTATION IN NATURALLY REDUCTIVE HOMOGENEOUS SPACES

Ricardo Ferreira, João Xavier, João Paulo Costeira, Victor Barroso

Instituto de Sistemas e Robótica - Instituto Superior Técnico
Laboratório de Processamento de Sinal
Av. Rovisco Pais, 1049-001 Lisboa
ricardo@isr.ist.utl.pt, jxavier@isr.ist.utl.pt, jpc@isr.ist.utl.pt, vab@isr.ist.utl.pt

ABSTRACT

We address the problem of computing the Riemannian centroid of a constellation of points in a naturally reductive homogeneous manifold. We note that many interesting manifolds used in engineering (such as the special orthogonal group, Grassman, sphere, positive definite matrices) possess this structure. We develop an intrinsic Newton scheme for the centroid computation. This is achieved by exploiting a formula that we introduce for obtaining the Hessian of the squared Riemannian distance on naturally reductive homogeneous spaces. Some results of finding the centroid of a constellation of points in these spaces are presented, which evidence the quadratic convergence of the Newton method derived herein. These computer simulation results show that, as expected, the Newton method has a faster convergence rate than the usual gradient-based approaches.

1. INTRODUCTION AND MOTIVATION

Due to its quadratic convergence rate near the solution, Newton's method has for a long time been the method of choice for optimization problems, especially when high precision is required, in all fields from engineering to numerical analysis where it is used extensively to obtain many digits of precision. Intrinsic Newton algorithms for optimization problems with orthogonality and unitary constraints are discused in [1] and [2]. The focus of this paper is on deriving an intrinsic Newton algorithm for computing the center of mass in naturally reductive homogeneous spaces.

1.1 Applications of Center of Mass Computation Moakher in his study of $SO(3)$ [3] mentions the study of plate tectonics and sequence-dependent continuum modeling of DNA where experimental observations are obtained with a significant amount of noise that needs to be smoothed. Manton confirms this need in [4] and further broadens the applications to fuzzy control, robotics and vision. Pennec [5] states that positive definite symmetric matrices are commonly used as covariance matrices for statistical characterization of deformations and encoding of principle diffusion directions in Diffusion Tensor Imaging (DTI), expanding the range of applications to medicine. Computation of centers of mass also find applications for analyzing shapes in medical imaging, see [6]. Center of mass computation is also a mandatory step when considering the extension of the K-means algorithm to manifolds.

It is important to note that all of the manifolds mentioned, and many other commonly used fall under the class of naturally reductive homogeneous spaces (see [7] for an introduction) considered in this paper. The Grassman manifold ($\mathbb{G}$), sphere ($\mathbb{S}$), positive definite matrices (S) and the special orthogonal group (SO) are only a subset of spaces that share this property.

1.2 State of the art Several approaches to the optimization of this cost function exist, most of them relying on gradient methods. For example Moakher has a study [3] of the problem in $SO(3)$ in which he presents solutions to particular cases. Manton [4] presents a gradient method applicable to compact Lie groups proven to be globally convergent as long as the points are close enough to each other. Hüper and Manton [8] developed a Newton method for this cost function on the orthogonal group. The algorithm presented by Pennec in [5] for the space of positive definite matrices is also a gradient descent method.

1.3 Problem Formulation Let M be a naturally reductive homogeneous space [7] and $\mathcal{X} = \{p_1, \ldots, p_L\} \subset M$ a constellation of P points. Let $d : M \times M \to \mathbb{R}$ be the function that returns the intrinsic distance of any two points on the manifold and define a cost function $C_{\mathcal{X}} : M \to \mathbb{R}$ as

$$C_{\mathcal{X}}(q) = \frac{1}{2} \sum_{l=1}^{L} d(p_l, q)^2 = \sum_{l=1}^{P} k_{p_l}(q) \ , \qquad (1)$$

where the functions $k_{p_l} : M \to \mathbb{R}$ consider the distance to each point individually and are defined as $k_{p_l}(q) = 1/2 \ d(p_l, q)^2$. The Fréchet mean set of the constellation is defined as the set of solutions to the optimization problem $m_f(\mathcal{X}) = \operatorname{argmin}_{q \in M} C_{\mathcal{X}}(q)$. Each element of the set $m_f(\mathcal{X})$ will be called a centroid of $\mathcal{X}$. Note that depending on the manifold M a generic constellation might have more than one centroid (for example if the sphere is considered with a constellation consisting of two antipodal points, all the equator points are centroids). The set of points at which the function (1) attains a local minimum is called the Karcher mean set and is denoted as $m_k(\mathcal{X})$. The objective will be to find a centroid for the given constellation (which in the applications of interest should be unique), but the possibility of convergence to a local minimum is not dealt with. If the points on the constellation are close enough to each other, it is known that the global set $m_f(\mathcal{X})$ has a single element and so the centroid is unique as stated in [4] and [9].

1.4 Contribution The contribution of this paper is the construction of a Newton algorithm to compute centroids of constellations in naturally reductive homogeneous spaces, henceforth denoted by NRHS. These manifolds are created as a quotient space of a Lie group G by a closed Lie subgroup $H \subset G$ with certain properties (see [7] for an introduction on these manifolds). This paper focuses on the case $G = \mathbb{GL}(n, \mathbb{R})$, the set of invertible matrices. We achieve the algorithm construction by deriving a formula for the intrinsic Hessian

of the cost function considered in (1), which holds for this category of smooth manifolds. In fact, our result is valid for a wider range of manifolds since they only need to be locally symmetric Riemannian manifolds (see [7] [10] and [11] for an introduction). However, the need for carrying out some intrinsic computations such as parallel transport of vectors and Riemannian log maps, restricts the immediate feasibility of the approach mainly to NRHS.

1.5 Paper Organization Section 2 starts by reviewing Newton's method on Riemannian manifolds and then presents our theorem, needed for actual computation of the Hessian matrix for our case. Sections 3-5 illustrates the application of our Newton method for particular scenarios: we consider examples in the unit sphere, $SO(n)$ and $S(n)$. The performance of our algorithm is compared to the usual gradient-based approaches and shown to outperform them. Finally conclusions are drawn in section 6.

2. HESSIAN OF THE RIEMANNIAN SQUARED DISTANCE IN NATURALLY REDUCTIVE HOMOGENEOUS SPACES

2.1 Review of Newton's Method in Riemannian Manifolds Let $q_k \in M$ henceforth designate the kth iterate in an optimization method. Newton's method on a manifold is essentially the same as in $\mathbb{R}^n$ (see [1], [2] and [12] for some generalizations). It generates a search direction $d_k \in T_{q_k}M$ as the solution of the linear system

$$H \cdot d_k = -\operatorname{grad} f(q_k) \ , \tag{2}$$

where H is the matrix representation of the Hessian of the cost function and $\operatorname{grad} f(q_k) \in T_{q_k}M$ is its gradient. Some care is needed though, since the Hessian and the gradient are not as simple to find as in $\mathbb{R}^n$, but are in fact given as the solutions of $(df)_q X_q = \langle \operatorname{grad} f(q), X_q \rangle$ and $\operatorname{Hess} f(q)(X_q, Y_q) = \left\langle \nabla_{X_q} \operatorname{grad} f, Y_q \right\rangle$ for $q \in M$ and $X_q, Y_q \in T_qM$ are any tangent vectors, $(df)_q$ denotes the differential of the function f at the point q and ∇ denotes the Levi-Civita connection of the manifold.

Once a Newton direction has been obtained, it should be checked if it's a descent direction (its dot product with the gradient vector should be negative). If so, the update equation $q_{k+1} = exp_{q_k}(\alpha_k d_k)$, can be used to obtain a better estimate, where α_k is a step size, given for example by Armijo's rule and the Riemannian exp map provides a means of travel on the manifold (it shall be described in section 2.4). If the dot product is negative, a standard negative gradient direction should be used.

Using linearity of the gradient and the Hessian, the cost function in equation (1) allows for the decomposition

$$\begin{aligned} \operatorname{grad} C_{\mathcal{X}}(q) &= \sum_{l=1}^{L} \operatorname{grad} k_{p_l}(q) = -\sum_{l=1}^{L} log_q(p_n) \\ \operatorname{Hess} C_{\mathcal{X}}(q) &= \sum_{l=1}^{L} \operatorname{Hess} k_{p_l}(q) \ , \end{aligned} \tag{3}$$

where the Riemannian log map was used (it will be described in section 2.4). Although the gradient is readily computed (it is easy to check the result using normal coordinates), determination of the Hessian is more involved. The next section describes how to calculate it.

2.2 Calculating the Hessian

Theorem 2.1 *Consider a locally-symmetric n-dimensional Riemannian manifold M with curvature endomorphism R. Let $B_\epsilon(p)$ be a geodesic ball centered at $p \in M$ and $d_p : B_\epsilon(p) \to \mathbb{R}$ the function returning the distance from p. Let $\gamma : [0, r] \to B_\epsilon(p)$ denote the unit speed geodesic connecting p to $q \in B_\epsilon(p)$, where $r = d(p, q)$. Define the function $k_p : B_\epsilon(p) \to \mathbb{R}, k_p(q) = \frac{1}{2} d_p(q)^2$ and consider any $X_q, Y_q \in T_qM$. Then*

$$\begin{aligned} \operatorname{Hess}(k_p)_q(X_q, Y_q) = &\left\langle X_q^{\parallel}, Y_q^{\parallel} \right\rangle + \\ &+ r \sum_{i=1}^{n} \mathrm{c}_{\lambda_i}(r) \left\langle X_q^{\perp}, E_q^i \right\rangle \left\langle Y_q^{\perp}, E_q^i \right\rangle \ . \end{aligned} \tag{4}$$

where E_q^i is the parallel transport along γ of an orthonormal basis $E_p^i \in T_pM$ which diagonalizes the linear operator $\mathcal{R} : T_pM \to T_pM$, $\mathcal{R}(X_p) = R(X_p, \dot\gamma(0))\dot\gamma(0)$ with eigenvalues λ_i, this means $\mathcal{R}(E_p^i) = \lambda_i E_p^i$. Also,

$$\mathrm{c}_\lambda(t) = \begin{cases} \sqrt{-\lambda}/\tanh(\sqrt{-\lambda}\, t) & \lambda < 0 \\ \frac{1}{t} & \lambda = 0 \\ \sqrt{\lambda}/\tan(\sqrt{\lambda}\, t) & \lambda > 0 \end{cases} \ .$$

Here the $\parallel$ and $\perp$ signs denote parallel and perpendicular orthogonal components of the vector with respect to the velocity vector of γ, i.e. $X_q = X_q^{\parallel} + X_q^{\perp}$, $\left\langle X_q^{\perp}, X_q^{\parallel} \right\rangle = 0$, and $\left\langle X_q^{\perp}, \dot\gamma(d) \right\rangle = 0$.

Due to paper length constraints, the proof of this theorem is not presented. It can be found in [13]. The proof is established by exploiting some results about Jacobi fields [9] together with the fact that the curvature endomorphism R is parallel in these manifolds.

2.3 Algorithm skeleton

Input: Constellation $\mathcal{X} = \{p_1, ..., p_L\} \in M$
Output: Karcher Mean $q \in m_k(\mathcal{X})$
Initialization: choose $q_0 \in M$ and tolerance $\epsilon > 0$. Set $k = 0$.

Loop:
- Compute intrinsic gradient $g_k = \operatorname{grad} f(q) \in T_{q_k}M$.
- if $|g_k| < \epsilon$ set $q = q_k$ and return.
- compute Newton direction d_k.
- if $\langle d_k, g_k \rangle \geq 0$ set $d_k = -g_k$.
- apply Armijo rule to obtain $\alpha_k \approx \operatorname{argmin}_{\alpha \geq 0} exp_{q_k}(\alpha d_k)$.
- set $q_{k+1} = exp_{q_k}(\alpha_k d_k)$. Please note that due to finite precision limitations, after a few iterations the result should be enforced to lie on the the manifold.
- set $k \leftarrow k + 1$ and re-run the loop.

2.4 Implementation Considerations To implement equation (4) some intrinsic manifold computations need to be carried out. The next sections, describing the application of the algorithm to a specific set of manifolds contain the actual functions. The exponential map sends a vector $X_p \in T_pM$ to a point on the manifold. If γ is the unique geodesic such that $\gamma(0) = p$ and $\dot\gamma(0) = X_p$, then $exp_p(X_p) = \gamma(1)$. In general exp_p is only defined on a neighborhood of the origin in T_pM, however a NRHS is complete, which means exp_p has domain T_pM. On a sufficiently small open neighborhood, this map is a diffeomorphism. Its inverse function known as the logarithm, when defined, returns $X_p = log_p(q)$ such that, $\gamma(0) = p$, $\gamma(1) = q$ and $\dot\gamma(0) = X_p$. There is another important function which parallel transports a given vector along a geodesic. So $par_p(X_p, Y_p) : T_pM \times T_pM \to T_qM$ parallel translates the vector Y_p along the geodesic with the same characteristics as before. We note that in NRHS manifolds these operations are easy to carry out, e.g. see the closed form solutions in sections 3-5.

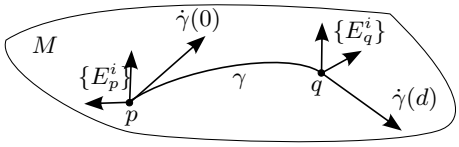

Fig. 1. Illustration of the orthonormal base needed.

A few remarks on how to obtain the orthonormal base set $\{E_q^i\}$ are in order. First let $\{F_p^i\} \subset T_pM$ be any orthonormal base and construct the symmetric matrix $A = [a_{ij}]$ where $a_{ij} = \left\langle F_p^i, \mathcal{R}(F_p^j)\right\rangle$. Let $A = VDV^T$ be its eigenvalue decomposition (EVD), where $D = \text{diag}(\lambda_1, ..., \lambda_n)$ reveals the necessary λ_i and $V = [v_{ij}]$ are used to construct $E_p^i = \sum_{k=1}^n v_{ki} F_p^k$. Finally, each element of this base is parallel translated to q resulting in the set $\{E_q^i\} \subset T_qM$. Please note that these λ_i are associated with the sectional curvature of the manifold, so constant sectional curvature spaces allow for an important simplification of equation (4):

$$\text{Hess}(k_p)_q(X_q, Y_q) = \left\langle X_q^{\parallel}, Y_q^{\parallel}\right\rangle + r\, \text{c}_\lambda(r) \left\langle X_q^{\perp}, Y_q^{\perp}\right\rangle \ , \quad (5)$$

where the constant λ is the sectional curvature. Note that the computational weight of calculating an EVD for every point in the constellation at every iteration of the Newton method is eliminated.

Whichever formula is used, (4) or (5) depending on the manifold considered, equation (3) characterizes the Hessian of the cost function considered from the individual Hessians, one for each point in the constellation.

Now that a means to obtain the value of the Hessian for any two tangent vectors is known, it is necessary to find its matrix representation. For that, consider any orthonormal base $\{E_q^i\} \subset T_qM$ and apply equation (4) or (5) to each pair of vectors, i.e. $H = [h_{ij}]$ where $h_{ij} = \text{Hess}\ (C_{\mathcal{X}})_q(E_q^i, E_q^j)$. Note that this matrix representation is valid only on this base, so it is necessary to write the gradient vector in this base when solving equation (2), i.e. $grad(f)_p = \sum_{i=1}^n g_i E_q^i$ where $g_i = \left\langle grad(f)_p, E_q^i\right\rangle$. Once the linear system $H \cdot d = -g$ is solved for d, the solution is once again expressed in this basis, so the Newton direction is $d_N = \sum_{i=1}^n d_i E_q^i$.

The following sections compare results obtained using our proposed Newton method and the standard gradient method commonly used (e.g. [5]) which iterates $q_{k+1} = exp_{q_k}(-\,\text{grad}\, C_{\mathcal{X}}(q_k))$.

Note that since we do not have access to the optimum, the last value given by the Newton algorithm is used as such. If the gradient algorithm converges to the same point, this is assumed to be the best estimate for the local minimizer.

3. $\mathbb{S}(N)$ - THE SPHERE

This n-dimensional manifold is described as the set $\mathbb{S}(n) = \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$ whose tangent space at a point $p \in \mathbb{S}(n)$ is $T_p\mathbb{S}(n) \cong \{x \in \mathbb{R}^{n+1} : p^T x = 0\}$. Let $p, q \in \mathbb{S}(n)$, $X_p, Y_p, Z_p \in T_p\mathbb{S}(n)$ and s is the norm of X_p. It can be shown that for the ambient metric $\langle X_p, Y_p\rangle = X_p^T Y_p$:

- $exp_p(X_p) = p\cos(s) + \frac{X_p}{s} sin(s)$.
- $log_p(q) = (q - p(p^T q))\frac{a}{\sin(a)}$ where $a = \arccos(p^T q)$.
- $par_p(X_p, Y_p) = -\langle X_p/s, Y_p\rangle (\sin(s)p + (1 - \cos(s))X_p/s) + Y_p$.
- $R(X_p, Y_p) \cdot Z_p = \langle Y_p, Z_p\rangle X_p - \langle X_p, Z_p\rangle Y_p$.

The constellation was populated with random points $p_l \in \mathbb{S}(n)$ generated as $p_l = \hat{p}_l/\|\hat{p}_l\|$ where $\hat{p}_l \in \mathbb{R}^{n+1}$ is randomly generated. In order to satisfy some distance constraints the last coordinate was enforced to be positive, i.e. all points were generated on the positive half sphere. Note that this is a constant sectional curvature space so equation (5) may be used to calculate the Hessian. The results obtained when considering $\mathbb{S}(4)$ with a constellation with 10 points are presented in figure 2. The Newton quadratic convergence rate is evident as is the linear rate associated with the gradient method.

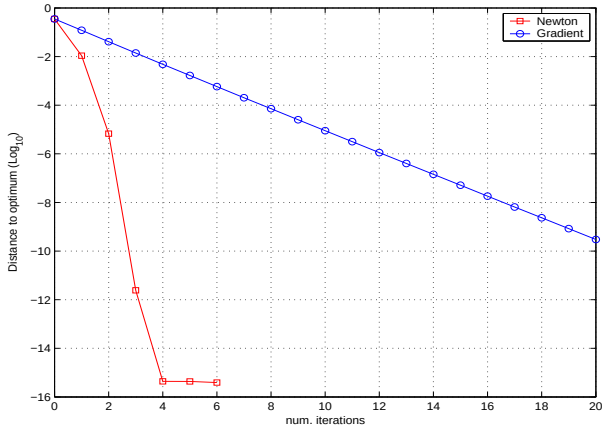

Fig. 2. Distance to optimum value in $\mathbb{S}(4) \subset \mathbb{R}^5$ with 10 random generated points.

4. $SO(N)$ - SPECIAL ORTHOGONAL GROUP

This $n(n-1)/2$ dimensional manifold represents the set of rotations of $\mathbb{R}^n$ and is described as $SO(n) = \{x \in M_n(\mathbb{R}) : x^T x = id\}$ ($M_n(\mathbb{R})$ is the set of $n \times n$ matrices with real entries) whose tangent space at a point $p \in SO(n)$ is $T_pSO(n) \cong \{pk : k \in \mathcal{K}(n, \mathbb{R})\}$, where $\mathcal{K}(n, \mathbb{R})$ denotes the set of $n \times n$ skew-symmetric matrices.

The constellation points are generated using a `QR` decomposition of a random matrix, guaranteeing that the result restricts to $SO(n)$ by checking the sign of the determinant. No other constraints were enforced so the points should spread through the whole manifold and not restrict to a particular neighborhood. Note that although $SO(3)$ has constant curvature, the same does not apply for higher order manifolds so the general equation (4) has to be used. The results shown in figure 3 were obtained using a constellation of 10 points on $SO(4)$. Note that this is a higher dimensional manifold than the previous sphere example (6 degrees of freedom compared to 4). Let $p, q \in SO(n)$, $X_p, Y_p, Z_p \in T_pSO(n)$. We have the following closed-form expressions for the ambient metric $\langle X_p, Y_p\rangle = \text{tr}\{X_p^T Y_p\}$:

- $exp_p(X_p) = p\exp(p^T X_p)$, where exp denotes the matrix exponential function.
- $log_p(q) = p\log(p^T q)$ where log denotes the matrix logarithm.
- $par_p(X_p, Y_p) = p\exp(p^T X_p/2)p^T Y_p \exp(p^T X_p/2)$.
- $R(X_p, Y_p) \cdot Z_p = -\frac{1}{4}\left[\left[X_p, Y_p\right], Z_p\right]$ where the brackets denote the lie bracket at point p.

Again, it is clear that the Newton method is converging at a quadratic rate, as opposed to the linear rate of the gradient method.

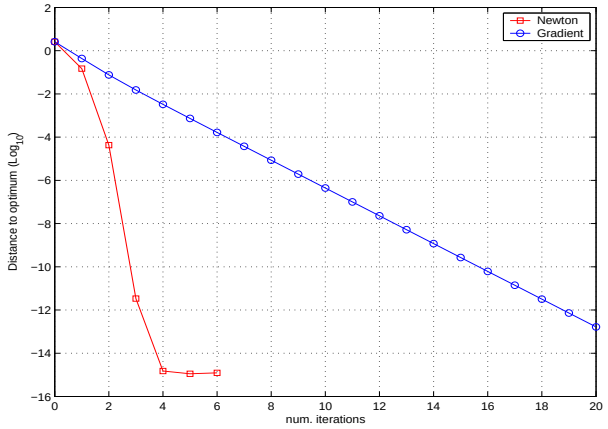

Fig. 3. Distance to optimum value in $SO(4) \subset M_4(\mathbb{R})$ with 10 random generated points.

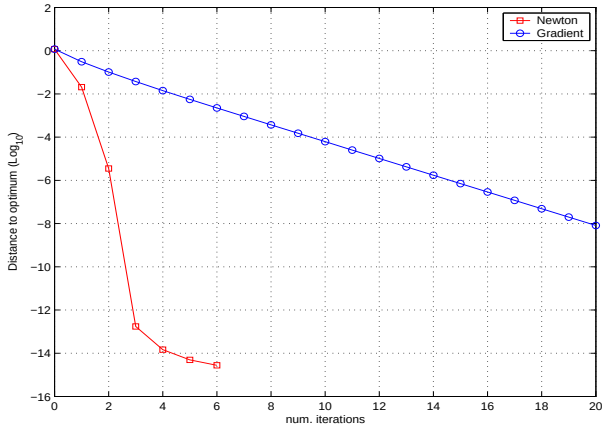

Fig. 4. Distance to optimum value in $S(3) \subset M_3(\mathbb{R})$ with 30 random generated points.

5. $S(N)$ - SYMMETRIC POSITIVE DEFINITE MATRICES

This $n(n+1)/2$ dimensional manifold is described as the set $S(n) = \{x \in M_n(\mathbb{R}) : x = x^T$, with positive eigenvalues$\}$ whose tangent space at a point $p \in S(n)$ is $T_pS(n) \cong \{x : x \in \mathcal{S}(n, \mathbb{R})\}$, where $\mathcal{S}(n, \mathbb{R})$ denotes the set of $n \times n$ symmetric matrices. Let $p, q \in S(n)$, $X_p, Y_p, Z_p \in T_pS(n)$. When considering the metric $\langle X_p, Y_p \rangle = \mathrm{tr}\{X_p^T p^{-1} Y_p p^{-1}\}$ we have, see [13]

- $exp_p(X_p) = p^{1/2} \exp(p^{-1/2} X_p p^{-1/2}) p^{1/2}$.
- $log_p(q) = p^{1/2} \log(p^{-1/2} q p^{-1/2}) p^{1/2}$.
- $par_p(X_p, Y_p) = p^{1/2} Q p^{-1/2} Y_p p^{-1/2} Q p^{1/2}$, where $Q = \exp(p^{-1/2} X_p p^{-1/2}/2)$.
- $R(X_p, Y_p) \cdot Z_p = 1/4(Z_p p^{-1} O - O p^{-1} Z_p)$, where $O = X_p p^{-1} Y_p - Y_p p^{-1} X_p$.

In this example 30 points on $S(3)$ (a 6 dimensional manifold) were generated to form a constellation to which the algorithm has been applied. This manifold has the particularity of having non-positive sectional curvature, guaranteeing the uniqueness of the centroid. The points were generated around a nominal point by following geodesics departing from this point with tangent vectors obeying a Gaussian distribution. This constellation construction method and the manifold considered resembles the one described in [5]. The results are shown in figure 4. Once again the quadratic convergence rate of the Newton method outperforms the gradient method.

6. CONCLUSIONS

A way to compute the Hessian of the squared distance cost function on naturally reductive homogeneous spaces was presented and some results of its application to calculate the Karcher mean of a set of points on commonly used manifolds were shown. The results clearly show that when used in a Newton optimization algorithm, if it converges, it does so quadratically fast. Comparing with standard gradient algorithms this results in a faster convergence rate, requiring in some cases only a small fraction of the iterations for an adequate solution. For constant sectional curvature spaces the computational cost per Newton iteration affords a significant reduction. In [13] we study some approximations to the Hessian with the intent of relieving the computational cost per Newton iteration.

7. REFERENCES

[1] Alan Edelman, Tomas A. Arias, and Steven T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1999.

[2] Jonathan H. Manton, "Optimisation algorithms exploiting unitary constraints," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 635–650, March 2002.

[3] Maher Moakher, "Means and averaging in the group of rotations," *SIAM J. Matrix Anal. Appl.*, vol. 24, no. 1, pp. 1–16, 2002.

[4] Jonathan H. Manton, "A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups," in *Eighth International Conference on Control, Automation, Robotics and Vision*, Kunming, China, December 2004.

[5] Xavier Pennec, Pierre Fillard, and Nicholas Ayache, "A Riemannian framework for tensor computing," Research Report 5255, INRIA, July 2004, To appear to the Int. Journal of Computer Vision.

[6] P. Fletcher, C., Lu S.Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Transactions on Medical Imaging*, 2004.

[7] Barret O'Neil, *Semi-Riemannian Geometry*, Academic Press, 1983.

[8] K. Hüper and J. Manton, "The Karcher mean of points on the orthogonal group," *Preprint*, 2005.

[9] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Comm. Pure Appl. Math.*, vol. 30, pp. 509–541, 1977.

[10] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, 1975.

[11] Manfredo P. do Carmo, *Riemannian Geometry*, Birkhäuser, Boston, MA, 1992.

[12] Knut Hüper and Jochen Trumpf, "Newton-like methods for numerical optimization on manifolds," *Asilomar Conference on Signals, Systems and Computers*, pp. 136–139, 2004.

[13] R. Ferreira, J. Xavier, J. Costeira, and V. Barroso, "Intrinsic newton method for computing Riemannian centroids in naturally reductive homogeneous spaces," *In Preparation*.