On the computation of means on Grassmann manifolds

Knut Hüper, Uwe Helmke and Sven Herzberg

***Abstract*— Given a set of data points on a Grassmann manifold sufficiently close to each other, one way to define their centroid or geometric mean is via the minimizer of a certain cost function. If one chooses the cost as the sum of squared geodesic distances between a given point and all the data points we end up with the definition of the Karcher mean. In this paper we analyze the critical points for this cost function.**

I. INTRODUCTION

In recent years numerical methods for the computation of means on differentiable manifolds have awoken increased interest. Since the seminal paper by Karcher, *cf.* [13], several papers have been published, emphasizing mainly differential geometric aspects as well as statistical applications in connection with convexity concepts, *cf.* [8], [14], [15], to cite just a few. The reason for new interest into the topic of computing averages, means or the like on differentiable manifolds, lies in the fact, that in applications, noise is often an intrinsic obstacle during the process of measurement. In modern engineering applications state variables as well as other observables do not lie necessarily in a vector space. Instead, they might be elements of a nonlinear differentiable manifold. This often requires the generalization of the concept of means and other statistical tools, now also from a computational point of view.

There has been done already a lot in this direction. We mention computer graphics, *cf.* [7], pose estimation in robotics, DNA-modeling or geology, where averaging over rotations play an important role, *cf.* [9], [19], linear algebra and matrix analysis, *cf.* [3], [4], [5], [6], [20], signal processing, *cf.* [18], variational problems, *cf.* [16], and more recently also in medical imaging, computer vision and distributed consensus problems [21], [22], [23].

In this paper we discuss the computation of the Karcher mean on the real Grassmann manifold. Only a little has been done in this direction for this important compact symmetric space, see however [10] or [2]. The Riemannian geometry of this manifold is exploited to derive explicit formulas for the exponential map and its inverse. The critical point set of the distance function related to the Karcher mean is computed. To derive a gradient or Newton-like algorithm is then straight forward.

II. THE GRASSMANN MANIFOLD

Recall, that the Grassmann manifold $\mathrm{Gr}_{m,n}$ is defined as the set of m-dimensional $\mathbb{R}$-linear subspaces of $\mathbb{R}^n$. It is a smooth, compact manifold and provides a natural generalization of the familiar projective spaces. Thus we define the *Grassmannian* as

$$\mathrm{Gr}_{m,n} := \{P \in \mathbb{R}^{n\times n} \mid P^\top = P, P^2 = P, \operatorname{tr} P = m\}, \quad (1)$$

the manifold of rank m symmetric projection operators of $\mathbb{R}^n$; see [11] for the construction of a natural bijection with the Graßmann manifold and a proof that it defines a diffeomeorphism. Let

$$\mathrm{Sym}_n := \{S \in \mathbb{R}^{n\times n} \mid S^\top = S\} \quad (2)$$

and

$$\mathfrak{so}_n := \{\Omega \in \mathbb{R}^{n\times n} \mid \Omega^\top = -\Omega\} \quad (3)$$

denote the vector spaces of real symmetric and real skew-symmetric matrices, respectively.

Theorem 1: (a) The Grassmannian $\mathrm{Gr}_{m,n}$ is a smooth, compact submanifold of Sym_n of dimension $m(n-m)$.

(b) The tangent space of $\mathrm{Gr}_{m,n}$ at an element $P \in \mathrm{Gr}_{m,n}$ is given as

$$T_P \,\mathrm{Gr}_{m,n} = \{[P,\Omega] \mid \Omega \in \mathfrak{so}_n\}. \quad (4)$$

Here $[P,\Omega] := P\Omega - \Omega P$ denotes the matrix commutator (Lie bracket).

For any $P \in \mathbb{R}^{n\times n}$ let

$$\mathrm{ad}_P : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}, \quad \mathrm{ad}_P(X) := [P,X]. \quad (5)$$

There are at least two natural Riemannian metrics defined on the Grassmannian $\mathrm{Gr}_{m,n}$, the induced *Euclidean metric* and the *normal metric*, cf. e.g. [11] or [17].

The Euclidean Riemannian metric on $\mathrm{Gr}_{m,n}$ is defined by the Frobenius inner product on the tangent spaces

$$\langle X, Y\rangle := \operatorname{tr}(XY) \quad (6)$$

for all $X, Y \in T_P \,\mathrm{Gr}_{m,n}$ which is induced by the embedding space Sym_n.

The normal Riemannian metric has a somewhat more complicated definition. Consider the surjective linear map

$$\mathrm{ad}_P : \mathfrak{so}_n \to T_P \,\mathrm{Gr}_{m,n}, \quad \Omega \mapsto [P,\Omega] = \mathrm{ad}_P\,\Omega \quad (7)$$

with kernel

$$\ker \mathrm{ad}_P = \{\Omega \in \mathfrak{so}_n \mid P\Omega = \Omega P\}. \quad (8)$$

We regard $\mathfrak{so}_n$ as an inner product space, endowed with the Frobenius inner product $\langle \Omega_1, \Omega_2\rangle = \operatorname{tr}(\Omega_1^\top \Omega_2) = -\operatorname{tr}(\Omega_1\Omega_2)$. Then ad_P induces an isomorphism of vector spaces

$$\widehat{\mathrm{ad}}_P : (\ker \mathrm{ad}_P)^\perp \to T_P \,\mathrm{Gr}_{m,n} \quad (9)$$

The authors are with Institut für Mathematik, Julius-Maximilians-Universität Würzburg, Am Hubland, 97074 Würzburg, Germany {hueper,helmke}@mathematik.uni-wuerzburg.de

ISBN 978-963-311-370-7

and therefore induces an isometry of inner product spaces, by defining an inner product on $T_P\,\mathrm{Gr}_{m,n}$ via

$$\langle\langle X, Y\rangle\rangle_P := -\operatorname{tr}(\widehat{\operatorname{ad}}_P^{-1}(X)\widehat{\operatorname{ad}}_P^{-1}(Y)). \tag{10}$$

Note, that this inner product on $T_P\,\mathrm{Gr}_{m,n}$, called the normal Riemannian metric, might vary with the basepoint P. Luckily, the situation is better than one would expect, as Proposition 1 below shows.

Proposition 1: The Euclidean and normal Riemannian metrics on the Grassmannian $\mathrm{Gr}_{m,n}$ coincide, i.e. for all $P \in \mathrm{Gr}_{m,n}$ and for all $X, Y \in T_P\,\mathrm{Gr}_{m,n}$ we have

$$\operatorname{tr}(X^\top Y) = -\operatorname{tr}\left(\widehat{\operatorname{ad}}_P^{-1}(X)\ \widehat{\operatorname{ad}}_P^{-1}(Y)\right). \tag{11}$$

Since these two Riemannian metrics on the Grassmannian coincide, they also define the same geodesics. Thus, in the sequel, we focus on the Euclidean metric. Note, that the above result is not true for arbitrary flag manifolds and in fact, the geodesics are then different for the two metrics. The following result characterizes the geodesics on $\mathrm{Gr}_{m,n}$.

Theorem 2: The geodesics of $\mathrm{Gr}_{m,n}$ are exactly the solutions of the second order differential equation

$$\ddot{P} + [\dot{P}, [\dot{P}, P]] = 0. \tag{12}$$

The unique geodesic $P(t)$ with initial conditions $P(0) = P_0 \in \mathrm{Gr}_{m,n}$, $\dot{P}(0) = \dot{P}_0 \in T_{P_0}\,\mathrm{Gr}_{m,n}$ is given by

$$P(t) = \mathrm{e}^{t[\dot{P}_0, P_0]}\, P_0\, \mathrm{e}^{-t[\dot{P}_0, P_0]}. \tag{13}$$

III. Geodesic distance on the Grassmannian

The above explicit formula for geodesics leads to the following formula for the geodesic distance between two points on a Grassmannian. We omit the simple proof; see also [1] for a slightly different formula which is only valid on an open and dense subset of the Grassmannian.

Corollary 1: Let $P, Q \in \mathrm{Gr}_{m,n}$. Given any $\Theta \in \mathrm{SO}_n$ such that

$$P = \Theta^\top \left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right] \Theta.$$

We might choose Θ such that

$$\left[\begin{smallmatrix} \Lambda & A \\ A^\top & \Sigma \end{smallmatrix}\right] := \Theta Q \Theta^\top$$

with diagonal Λ and diagonal Σ. Let $1 \geq \lambda_1 \geq \cdots \geq \lambda_m \geq 0$ denote the eigenvalues of Λ. The squared geodesic distance of P to Q in $\mathrm{Gr}_{m,n}$ is given by

$$\operatorname{dist}^2(P, Q) = 2\textstyle\sum_{i=1}^{m} \arccos^2(\sqrt{\lambda_i}). \tag{14}$$

Alternatively, let $1 \geq \sigma_1 \geq \cdots \geq \sigma_{n-m} \geq 0$ denote the eigenvalues of Σ. Then

$$\operatorname{dist}^2(P, Q) = 2\textstyle\sum_{i=1}^{n-m} \arcsin^2(\sqrt{\sigma_i}). \tag{15}$$

In particular, if $P, Q \in \mathrm{Gr}_{m,n}$ with $Q = YY^\top, Y^\top Y = I_m$, then

$$\operatorname{dist}^2(P, Q) = 2\operatorname{tr}\left(\arccos^2((Y^\top P Y)^{\frac{1}{2}})\right). \tag{16}$$

Note that formula (15) is more efficient in the case $2m > n$. Note also that our formulas imply that the maximal length of a simple closed geodesic in $\mathrm{Gr}_{m,n}$ is $\sqrt{2m} \cdot \pi$ for $2m \leq n$ and $\sqrt{2(n-m)} \cdot \pi$ for $2m > n$. Unfortunately, none of the formulas (14)-(16) is very well suited for a discussion of critical points.

IV. Critical points of the geodesic distance

Let $P, Q \in \mathrm{Gr}_{m,n}$ and assume for the moment that $Q = \left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]$. Assume further that P is sufficiently close to Q, i.e., that there exists a ***unique*** $Z = Z(P) \in \mathbb{R}^{(n-m)\times n}$ such that

$$P = \exp\left(\operatorname{ad}_{\left[\begin{smallmatrix} 0 & -Z^\top \\ Z & 0 \end{smallmatrix}\right]}\right)\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]. \tag{17}$$

For simplicity, we consider first

$$f\colon\ \mathrm{Gr}_{n,m} \to \mathbb{R}, \quad P \mapsto \operatorname{tr}\left[\begin{smallmatrix} 0 & Z^\top \\ Z & 0 \end{smallmatrix}\right]^2. \tag{18}$$

Note that

$$\begin{aligned} f(P) = \operatorname{dist}^2(P, Q) &= -\operatorname{tr}\left[\begin{smallmatrix} 0 & -Z^\top \\ Z & 0 \end{smallmatrix}\right]^2 = 2\operatorname{tr}(Z^\top Z) \\ &= -\operatorname{tr}\left(\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]\operatorname{ad}^2_{\left[\begin{smallmatrix} 0 & -Z^\top \\ Z & 0 \end{smallmatrix}\right]}\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]\right), \end{aligned} \tag{19}$$

where Z has to be considered as the unique solution of (17), i.e. as a function of P. Let

$$\mathcal{Z} := \left[\begin{smallmatrix} 0 & -Z^\top \\ Z & 0 \end{smallmatrix}\right]. \tag{20}$$

In the sequel we will use for the directional derivative the abbreviation $Z' := \mathrm{D}\, Z(P)P'$, with $P' \in T_P\,\mathrm{Gr}_{m,n}$ and we use $\mathcal{Z}'$ accordingly. Consequently,

$$\begin{aligned} \mathrm{D}\, f(P)P' &= 4\operatorname{tr} Z^\top Z' = -2\operatorname{tr}(\mathcal{Z}\mathcal{Z}') \\ &= -2\operatorname{tr}\left(\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]\operatorname{ad}_{\mathcal{Z}}\operatorname{ad}_{\mathcal{Z}'}\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]\right). \end{aligned} \tag{21}$$

We need an expression for $\operatorname{ad}_{\mathcal{Z}'}$. Taking the derivative of (17) with respect to P in direction P' gives

$$P' = \mathrm{e}^{\operatorname{ad}_{\mathcal{Z}}}\left(\left(\frac{\mathrm{id} - \mathrm{e}^{-\operatorname{ad}_{\operatorname{ad}_{\mathcal{Z}}}}}{\operatorname{ad}_{\operatorname{ad}_{\mathcal{Z}}}}\operatorname{ad}_{\mathcal{Z}'}\right)\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]\right). \tag{22}$$

By parallel transporting $\left[\begin{smallmatrix} 0 & K^\top \\ K & 0 \end{smallmatrix}\right] \in T_Q\,\mathrm{Gr}_{m,n}$ along the unique geodesic connecting Q and P with respect to the Riemannian connection, cf. [12], it is easily seen that $P' \in T_P\,\mathrm{Gr}_{m,n}$ has the representation

$$P' = \mathrm{e}^{\operatorname{ad}_{\mathcal{Z}}}\left[\begin{smallmatrix} 0 & K^\top \\ K & 0 \end{smallmatrix}\right] \tag{23}$$

with $K \in \mathbb{R}^{(n-m)\times m}$. Using

$$\mathcal{K} := \left[\begin{smallmatrix} 0 & -K^\top \\ K & 0 \end{smallmatrix}\right] \tag{24}$$

and (23) a somewhat lengthy computation including the decomposition of the function $x \mapsto \frac{1-\mathrm{e}^{-x}}{x} = \frac{\sinh x}{x} + \frac{1-\cosh x}{x}$ into even and odd part shows that (22) is equivalent to

$$\left[\begin{smallmatrix} 0 & K^\top \\ K & 0 \end{smallmatrix}\right] = \operatorname{ad}_{\mathcal{K}}\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right] = \left(\frac{\sinh \operatorname{ad}_{\operatorname{ad}_{\mathcal{Z}}}}{\operatorname{ad}_{\operatorname{ad}_{\mathcal{Z}}}}\operatorname{ad}_{\mathcal{Z}'}\right)\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]. \tag{25}$$

By assumption on P being close enough to $Q = \left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]$, the selfadjoint operator $\frac{\sinh \operatorname{ad}_{\operatorname{ad}_{\mathcal{Z}}}}{\operatorname{ad}_{\operatorname{ad}_{\mathcal{Z}}}}$ is invertible. By exploiting the representation property of the ad-operator, i.e. $[\operatorname{ad}_X, \operatorname{ad}_Y] = \operatorname{ad}_{[X,Y]}$ as well as linearity we can conclude that (25) is equivalent to

$$\mathcal{Z}' = \left(\frac{\sinh \operatorname{ad}_{\mathcal{Z}}}{\operatorname{ad}_{\mathcal{Z}}}\right)^{-1}\mathcal{K}. \tag{26}$$

In summary,

$$\begin{aligned}
&\mathrm{D}\, f(P)P' \\
&= -2\,\mathrm{tr}\Big(\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right] \mathrm{ad}_{\mathcal{Z}}\, \mathrm{ad}_{\left(\frac{\sinh \mathrm{ad}_{\mathcal{Z}}}{\mathrm{ad}_{\mathcal{Z}}}\right)^{-1}\mathcal{K}} \left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]\Big) \\
&= 2\,\mathrm{tr}\left(\left(\mathrm{ad}_{\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]} \mathcal{Z}\right)\left(\mathrm{ad}_{\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]} \left(\left(\frac{\sinh \mathrm{ad}_{\mathcal{Z}}}{\mathrm{ad}_{\mathcal{Z}}}\right)^{-1}\mathcal{K}\right)\right)\right) \\
&= -2\,\mathrm{tr}\left(\Big(\underbrace{\mathrm{ad}^2_{\left[\begin{smallmatrix} I_m & 0 \\ 0 & 0 \end{smallmatrix}\right]} \mathcal{Z}}_{=\mathcal{Z}}\Big)\left(\left(\frac{\sinh \mathrm{ad}_{\mathcal{Z}}}{\mathrm{ad}_{\mathcal{Z}}}\right)^{-1}\mathcal{K}\right)\right) \\
&= -2\,\mathrm{tr}\left(\mathcal{Z}\left(\frac{\sinh \mathrm{ad}_{\mathcal{Z}}}{\mathrm{ad}_{\mathcal{Z}}}\right)^{-1}\mathcal{K}\right).
\end{aligned} \tag{27}$$

Therefore, by the self-adjointness of the operator $\left(\frac{\sinh \mathrm{ad}_{\mathcal{Z}}}{\mathrm{ad}_{\mathcal{Z}}}\right)^{-1}$ and $x \mapsto \frac{x}{\sinh x}$ being an even function,

$$\begin{aligned}
\mathrm{D}\, f(P) = 0 &\iff \mathrm{tr}(Z^\top K) = 0 \quad \text{for all } K \in \mathbb{R}^{(n-m)\times m} \\
&\iff Z = 0.
\end{aligned} \tag{28}$$

Let $Q \in \mathrm{Gr}_{m,n}$ be arbitrary now and let P again be sufficiently close to Q. We can therefore express $P = \mathrm{e}^{[\xi(P),Q]}\, Q\, \mathrm{e}^{-[\xi(P),Q]}$ for unique $\xi \in T_Q \mathrm{Gr}_{m,n}$. Consider

$$f\colon\ \mathrm{Gr}_{m,n} \to \mathbb{R}, \quad P \mapsto \mathrm{tr}\,\xi\xi^\top = \mathrm{dist}^2(P,Q). \tag{29}$$

Then by invariance

$$\mathrm{D}\, f(P) = 0 \iff \xi = 0. \tag{30}$$

V. Means on the Grassmannian

We now consider a geodesically convex open ball $\mathcal{B} \subset \mathrm{Gr}_{m,n}$ containing several data points Q_i. Let $P \in \mathcal{B}$. Obviously, for each i there exists a unique $\xi_i \in T_{Q_i} \mathrm{Gr}_{m,n}$ with

$$P = \mathrm{e}^{\mathrm{ad}_{[\xi_i,Q_i]}}\, Q_i. \tag{31}$$

Let

$$\begin{aligned}
&f\colon \mathcal{B} \to \mathbb{R}, \\
&\quad P \mapsto \sum_i \mathrm{tr}\,\xi_i\xi_i^\top = \sum_i \mathrm{dist}^2(P,Q_i) \\
&\qquad\qquad\qquad = -\sum_i \mathrm{tr}\, Q_i\, \mathrm{ad}^2_{[\xi_i,Q_i]}\, Q_i.
\end{aligned} \tag{32}$$

As a straight forward generalization of the above results we get the well known fact that

$$\mathrm{D}\, f(P) = 0 \iff \sum_i \xi_i = 0. \tag{33}$$

The interpretation of this condition is easily understood. Let P be the unique critical point of f on $\mathcal{B}$. Attaching Riemannian normal coordinates around P tells us that in this chart the origin (i.e. P) is equal to the usual Euclidean geometric mean of all the data points Q_i, expressed in exactly this chart.

Theorem 3: The Riemannian gradient, with respect to the normal Riemannian (Euclidean) metric, of the function f, defined by (32), then is as

$$\mathrm{grad}\, f(P) = 2\sum_i \mathrm{ad}_{[\xi_i,Q_i]}\, P. \tag{34}$$

VI. Acknowledgement

This work was partially supported by the German Federal Ministry of Education and Research (BMBF) project FH-profUnd 2007, *Development and Implementation of Novel Mathematical Algorithms for Identification and Control of Technical Systems*.

References

[1] P.-A. Absil. *Invariant Subspace Computation: A Geometric Approach*. PhD thesis, University of Liege, Belgium, 2003.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80:199–220, 2004.

[3] T. Ando, C.-K. Li, and R. Mathias. Geometric means. *Linear Algebra Appl.*, 385:305–334, 2004.

[4] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.*, 29(1):328–347, 2007.

[5] R. Bhatia and J. Holbrook. Riemannian geometry and matrix geometric means. *Linear Algebra Appl.*, 413(2-3):594–618, 2006.

[6] S. Bonnabel and R. Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM. J. Matrix Anal. & Appl.*, 31(3), 2009.

[7] S. R. Buss and J. P. Fillmore. Spherical averages and applications to spherical splines and interpolation. *ACM Trans. Graph.*, 20(2):95–126, 2001.

[8] G. Galperin. A concept of the mass center of a system of material points in the constant curvature spaces. *Commun. Math. Phys.*, 154:63–84, 1993.

[9] C. Gramkow. On averaging rotations. *J. Math. Imaging Vis.*, 15(1-2):7–16, 2001.

[10] U. Helmke, K. Hüper, and J. Trumpf. Newton's method on Graßmann manifolds. arXiv:0709.2205v2, 2007.

[11] U. Helmke and J. Moore. *Optimization and dynamical systems. Foreword by R. W. Brockett.* Berlin: Springer, 1994.

[12] K. Hüper and F. Silva Leite. On the geometry of rolling and interpolation curves on S^n, SO_n, and Grassmann manifolds. *J. Dyn. Control Syst.*, 13(4):467–502, 2007.

[13] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Appl. Math.*, XXX:509–541, 1977.

[14] W. S. Kendall. Probability, convexity, and harmonic maps with small image. I: Uniqueness and fine existence. *Proc. Lond. Math. Soc., III. Ser.*, 61(2):371–406, 1990.

[15] H. Le. Locating Fréchet means with application to shape spaces. *Adv. Appl. Probab.*, 33(2):324–338, 2001.

[16] L. Machado, F. Silva Leite, and K. Hüper. Riemannian means as solutions of variational problems. *LMS J. Comput. Math.*, 9:86–103, March 2006.

[17] R. Mahony. *Optimization algorithms on homogeneous spaces*. PhD thesis, Australian National University, Canberra, March 1994.

[18] J. H. Manton. A centroid (Karcher mean) approach to the joint approximate diagonalisation problem: The real symmetric case. *Digital Signal Processing*, 16(5):468–478, 2006.

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

[20] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.*, 26(3):735–747, 2005.

[21] X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vis.*, 25(1):127–154, 2006.

[22] R. Subbarao and P. Meer. Nonlinear mean shift over Riemannian manifolds. *International Journal of Computer Vision*, 84:1–20, 2009.

[23] R. Tron, R. Vidal, and A. Terzis. Distributed pose averaging in camera sensor networks via consensus on SE(3). In *IEEE International Conference on Distributed Smart Cameras.*, 2008.