Rank-preserving geometric means of positive semi-definite matrices

Silvere Bonnabel Rodolphe Sepulchre

***Abstract*— The generalization of the usual geometric mean of two positive numbers a and b to positive definite matrices A and B has attracted considerable attention since the seminal work of Ando, and finds an increasing number of applications in signal and image processing. Building upon some recent work of the authors, the present paper proposes a generalization of any geometric mean defined on the interior of the cone of positive definite matrices, that is, for full rank matrices, to a rank-preserving geometric mean defined on the boundary of the cone, that is, for fixed-rank positive semidefinite matrices. The work is motivated by signal processing (e.g. filtering) operations on low-rank approximations of positive definite matrices in high-dimensional spaces. The paper will discuss the reasons why the proposed definition is sound and relevant in applications.**

***Index Terms*— Matrix means, geometric mean, positive semi-definite matrices, symmetries, singular value decomposition, principal angles.**

I. Introduction

Positive definite matrices have become fundamental computational objects in many areas of engineering and applied mathematics. They appear as covariance matrices in statistics, as variables in convex and semidefinite programming, as unknowns of important matrix (in)equalities in system and control theory, as kernels in machine learning, and as diffusion tensors in medical imaging, to cite a few. These applications have motivated the development of a difference and differential calculus over positive definite matrices. As its most basic operation, this calculus requires the proper definition of a mean. In particular, much research has been devoted to generalizing the geometric mean $\sqrt{ab}$ of two positive numbers a and b to positive definite matrices (see for instance Chapter 4 in [4] for an expository and insightful treatment of the subject). The further extension of a geometric mean from two to an arbitrary number of positive definite matrices is an active current research area [2], [14]. It has been increasingly recognized that from a theoretical point of view [7] as well as in numerous applications [13], [8], [12], [11], [2], [3], [14], [6], [16], [17], matrix geometric means are to be preferred to their arithmetic counterparts for developing a calculus in the cone of positive definite matrices.

The fundamental and axiomatic approach of Ando [2] (see also [14]) reserves the adjective "geometric" to a definition of mean that enjoys all the following properties:

- (P1) Consistency with scalars: if A, B commute $M(A,B) = (AB)^{1/2}$.
- (P2) Joint homogeneity $M(\alpha A, \beta B) = (\alpha\beta)^{1/2} M(A,B)$.
- (P3) Permutation invariance $M(A,B) = M(B,A)$.
- (P4) Monotonicity. If $A \leq A_0$ (i.e. $(A_0 - A)$ is a positive matrix) and $B \leq B_0$, the means are comparable and verify $M(A,B) \leq M(A_0, B_0)$.
- (P5) Continuity from above. If $\{A_n\}$ and $\{B_n\}$ are monotonic decreasing sequence (in the Lowner matrix ordering) converging to A, B then $\lim(A_n \circ B_n) = M(A,B)$.
- (P6) Congruence invariance. For any $G \in \mathrm{Gl(n)}$ we have $M(GAG^T, GBG^T) = GM(A,B)G^T$.
- (P7) Self-duality $M(A,B)^{-1} = M(A^{-1}, B^{-1})$.

The present paper seeks to extend any geometric mean defined on the open cone $\mathrm{P_n}$ to the the set of positive semi-definite matrices of fixed rank p, denoted by $S_+(p,n)$, which lies on the boundary of $\mathrm{P_n}$. Our motivation is primarily computational: with the growing use of low-rank approximations of matrices as a way to retain tractability in large-scale applications, there is an increasing need to extend the calculus of positive definite matrices to their low-rank counterparts. The classical approach in the literature is to extend the definition of a mean from the interior of the cone to the boundary of the cone by a continuity argument. As a consequence, this topic has not received much attention. This approach has however serious limitations from a computational viewpoint because it is not rank-preserving. We depart from this approach by viewing a rank p positive semi-definite matrix as the projection of a positive definite matrix in a p-dimensional subspace. The proposed mean is rank-preserving, and it possesses all the properties listed above, except (P6) that is shown to be impossible to preserve. (P7) must of course also be adapted to the set of rank-deficient matrices. Both are replaced with:

- (P6') Congruence invariance. For $(\mu, P) \in \mathbb{R} \times \mathrm{O(n)}$ we have $(\mu P^T A \mu P) \circ (\mu P^T B \mu P) = \mu P^T (A \circ B) \mu P$.
- (P7') Self-duality $M(A,B)^\dagger = M(A^\dagger, B^\dagger)$, where $\dagger$ is the pseudo-inversion.

The structure of the paper is as follows: in Section 2, we review the theory of Ando in the cone of positive definite matrices and we illustrate the shortcomings of the continuity argument for a rank-preserving mean to be defined on the boundary of the cone. In Section 3, we define and characterize the properties of the proposed mean. In Section 4, we show that the proposed rank-preserving geometric mean is rooted in a riemannian geometry recently studied in [5] for the set $S_+(p,n)$. Section 5 illustrates the relevance of a rank-preserving mean in the context of filtering. All proofs

S. Bonnabel is with Centre de Robotique, Mathématiques et Systèmes, Mines ParisTech, Boulevard Saint-Michel 60, 75272 Paris, France. silvere.bonnabel@mines-paristech.fr

Rodolphe Sepulchre is with Departement of Electrical Engineering and Computer Science, University of Liège, 4000 Liège, Belgium. {g.meyer,r.sepulchre}@ulg.ac.be

ISBN 978-963-311-370-7

and notations have been regrouped in the Appendix.

II. Analysis pass: Ando's approach

For positive scalars, the homogeneity property (P2) implies $M(a,b) = aM(1,b/a) = af(b/a)$ with f a monotone increasing continuous function. In a non-commutative matrix setting, one can write

$$M(A,B) = A^{1/2} f(A^{-1/2} B A^{-1/2}) A^{1/2} \quad (1)$$

Several geometric means can be defined in this way (see [3] and the Kosaki means). The popular geometric mean of two full-rank matrices introduced by Ando [1], [15], [2] corresponds to the case $f(X) = X^{1/2}$, generalizing the scalar geometric mean. It writes

$$M(A,B) = A\#B = A^{1/2}(A^{-1/2} B A^{-1/2})^{1/2} A^{1/2}. \quad (2)$$

It satisfies all the propositions P1-P7 listed above. There are many equivalent definitions of the Ando geometric mean. To cite a few, $A\#B = S^T(D_A B_B)^{1/2} S$ where S is any invertible matrix that simultaneously diagonalizes A and B by congruence. $A\#B$ can also be considered as the solution of the extremal problem:

$$\max \left\{ X \succeq 0 \ : \ \begin{pmatrix} A & X \\ X & B \end{pmatrix} \succeq 0 \right\} \quad (3)$$

A geometric mean satisfying (1) is defined for positive definite matrices, that is, for elements in the interior of the cone of positive definite matrices. Rank-deficient matrices lie on the boundary of the cone. As a consequence, the natural idea to extend a geometric mean on the boundary is to use a continuity argument. The resulting mean satisfies all the properties above (P7 must be formulated using pseudo-inversion), but it is not rank-preserving. Indeed, let $A = \mathrm{diag}(4, \epsilon^2)$ and $B = \mathrm{diag}(\epsilon^2, 1)$ where the term $\epsilon << 1$. These two matrices belong to P_2, and (Ando) geometric mean is $A\#B = \mathrm{diag}(2\epsilon,\ \epsilon)$. In the limit (rank-deficient) situation $\epsilon \to 0$, the mean becomes the null matrix.

A rank-preserving mean thus requires a different approach. We seek to retain most of the properties (P1-P7), but we will see that P6 must be relaxed in order to define a rank-preserving mean. We choose to replace the group of linear applications $\mathrm{Gl(n)}$ with the smaller but meaningful group of rotations and scaling $\mathbb{R}_+^* \times O(n)$.

III. Rank-preserving mean of PSD matrices

A. *Definition of the mean*

To design a rank-preserving mean on $\mathrm{S^+(p,n)}$, a starting point is the subset of rank p projectors:

$$\{P \in \mathbb{R}^{n\times n} /\ P^T = P,\ P^2 = P,\ \mathrm{Tr}\,(P) = p\}, \quad (4)$$

which is in bijection with the Grassman manifold of p-dimensional subspaces. The Riemannian mean of projectors in $\mathrm{Gr}(p,n)$ is a rank-preserving rotation-invariant mean. It seems reasonable to seek a rank-preserving mean that reduces to the Riemmannian mean of subspaces in the particular case of projectors. In order to extend the definition to the full set $\mathrm{S^+(p,n)}$, we use the polar decomposition

$$A = UR^2U^T$$

of any matrix $A \in \mathrm{S^+(p,n)}$, with $(U, R^2) \in \mathrm{V_{n,p}} \times \mathrm{P_p}$. Take k matrices in $\mathrm{S^+(p,n)}$ and for each i let $A_i = U_i R_i^2 U_i^T$. Because the $U_i U_i^T$'s are projectors, a tempting extension of an arbitrary geometric mean M would be the rank-preserving mean:

$$A_1 \circ ... \circ A_k = WM(R_1^2, \cdots, R_k^2)W^T$$

where M is applied to full rank matrices on the cone $\mathrm{P_p}$, and WW^T is the projector associated to the Grassman mean of projectors $U_1U_1^T$,...,$U_kU_k^T$.

This mean is not well defined. Indeed the matrix decomposition $A_i = U_i R_i^2 U_i^T$ is defined up to an orthogonal transformation $U \mapsto UO$, $R^2 \mapsto O^T R^2 O$ with $O \in \mathrm{O(p)}$ since $A_i = U_i R_i^2 U_i^T = U_i O (O^T R_i^2 O) O^T U_i$. The Grassman mean is unchanged by the transformation but, in general, $M(R_1^2, R_2^2) \neq M(R_1^2, O^T R_2^2 O)$. To obtain a well-defined mean from the geometric mean M, we need to select "canonical" representatives on each fiber $O(p)R_i^2 O(p)^T$.

For each $1 \le i \le k$, consider the singular value decomposition (SVD)

$$U_i^T W = O_i \mathrm{diag}(\sigma_1, \cdots, \sigma_p)(O_i^W)^T, \quad (5)$$

with $1 \ge \sigma_1 \ge \cdots \ge \sigma_p \ge 0$ where $O_i, (O_i^W)^T \in O(p)$ and where the singular values σ_i are the cosines of the principal angles $0 \le \theta_1 \le .. \le \theta_p \le \pi/2$ between the two subspaces range(V_i) and range(W) [9].

Because $WO_i^W = U_i O_i \mathrm{diag}(\sigma_1, \cdots, \sigma_p)$, the bases WO_i^W and $U_i O_i$ can be viewed as two special linked bases of the mean subspace WW^T, and the subspace under consideration $U_i U_i^T$. Indeed the geodesic linking one to another in the Stiefel manifold is the shortest amongst all geodesics of Stiefel linking both subspaces (i.e. it represents a geodesic in Grassman). Let us thus express the matrix A_i using the basis $U_i O_i$:

$$A_i = U_i O_i (O_i^T R_i^2 O_i)(U_i O_i)^T$$

The representative $\tilde{R}_i^2 = O_i^T R_i^2 O_i$ is thus a special interesting representative. And the matrix

$$\tilde{A}_i = (WO_i^W)\tilde{R}_i^2 (WO_i^W)^T$$

is the flat ellipsoid A_i brought to the mean subspace along a geodesic in the Grassmann manifold, i.e. by a rotation of mimimal energy. Its matrix representation in the fixed basis W is obviously $W^T \tilde{A}_i W = O_i^W \tilde{R}_i^2 (O_i^W)^T =: Q_i^2$.

Any geometric mean M defined over positive definite matrices can be extended to $\mathrm{S^+(p,n)}$ the following way

$$A_1 \circ ... \circ A_k = W[M(Q_1^2, ..., Q_k^2]W^T \quad (6)$$

The corresponding intuition is simple. Matrices of $\mathrm{S^+(p,n)}$ can be viewed as flat ellipsoids. The polar decomposition is such that U defines the subspace whereas R^2 defines the ellipsoid in the basis defined by U. All ellipsoids are mapped to the mean subspace by a rotation of minimal energy (the path of shortest length in $V_{n,p}$ corresponds to the Grassman geodesic based on the principal angles). Then, all the ellipsoids are expressed in a common basis of this subspace, and M is used to compute their geometric mean inside the subspace.

Proposition 1: Let M be any geometric mean on P_p. On the set of rank p projectors, the mean (6) coincides with the Grassman Riemannian mean. On the other hand, when the matrices all have the same range, (6) coincides with the geometric mean induced by M on the common range subspace of dimension p.

Proposition 2: Suppose $A_1, ...A_k$ are such that the Grassman mean of their ranges is unique and for each i the principal angles defined by (5) are different from $\pi/2$. If M satisfies properties P1-P7 the mean (6) can be called "geometric" as it satisfies all the properties P1-P5, P6'-P7'.

B. A simple example

The simplest case is given by rank one matrices in two dimensions $S^+(1,2)$ (see also [5]). Every matrix $A \in S^+(1,2)$ writes

$$A = xx^T = ur^2u^T$$

where $u = (\cos(\theta), \sin(\theta)) \in \mathbb{S}^1 \subset \mathbb{R}^2$ is a unit vector and (r, θ) is the polar representation of x. Without loss of generality θ is equated to $\theta + j\pi$, $j \in \mathbb{Z}$ since x and $-x$ correspond to the same A. Take θ_A, θ_B associated to u_A and u_B such that $|\theta_A - \theta_B| \leq \pi$.

Returning to the example of Section 2, we compute the rank-preserving mean of $A = \mathrm{diag}(4, 0)$ and $B = \mathrm{diag}(0, 1)$ to obtain the rank-one mean:

$$A \circ B = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \sqrt{4} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

The case of $S^+(1,2)$ gives insight in the general case, and allows to prove why a rank-preserving mean can not satisfy property (P6).

Lemma 1: There is no rank-preserving mean on $S^+(1,2)$ satisfying (P6).

IV. Geometric pass: geometric mean as a Riemannian mean

Any positive definite matrix admits the factorization $X = YY^T$, $Y \in Gl(n)$, and the factorization is invariant by rotation $Y \mapsto YO$. As a consequence, the cone of positive definite matrices has a homogeneous representation $Gl(n)/O(p)$. The space is reductive [7] and admits a Gl(n)-invariant metric. If X_1, X_2 are two tangent vectors at $A \in P_n$, the metric is given by $g_A^{P_n}(D_1, D_2) = \mathrm{Tr}\left(D_1A^{-1}D_2A^{-1}\right)$. With this definition, a geodesic (path of shortest length) at arbitrary $A \in P_n$ is [11], [17]: $\gamma_A(tX) = A^{1/2}\exp(tA^{-1/2}XA^{-1/2})A^{1/2}$, $t > 0$, and the corresponding geodesic distance is

$$d_{P_n}(A, B) = d(A^{-1/2}BA^{-1/2}, I) = \|\log(A^{-1/2}BA^{-1/2})\|_F = \sqrt{\sum_k \log^2(\lambda_k)},$$

where λ_k are the generalized eigenvalues of the pencil $A - \lambda B$, i.e., the roots of $\det(AB^{-1} - \lambda I)$. The distance is invariant with respect to action by congruence of Gl(n) and matrix inversion.

The geodesic characterization provides a closed-form expression of the Riemannian (Karcher) mean of two matrices $A, B \in P_n$. The geodesic $A(t)$ linking A and B is

$$A(t) = \exp_A^{P_n}(tX) = A^{1/2}\exp(t\log(A^{-1/2}BA^{-1/2}))A^{1/2},$$

where $A^{-1/2}XA^{-1/2} = \log(A^{-1/2}BA^{-1/2}) \in \mathrm{Sym(n)}$. The midpoint is obtained for $t = 1/2$: $M(A, B) = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$ and it corresponds to the Ando mean $A\#B$ [1].

In the recent work [5], we propose an extension of the affine-invariant metric of the cone to $S^+(p,n)$. If $(U, R^2) \in V_{n,p} \times P_p$ represents $A \in S^+(p,n)$, the tangent vectors of $T_AS^+(p,n)$ are represented by the infinitesimal variation (Δ, D), where

$$\begin{aligned} \Delta &= U_\perp B, \qquad B \in \mathbb{R}^{(n-p)\times p}, \\ D &= RD_0R \end{aligned} \tag{7}$$

such that $U_\perp \in V_{n,n-p}$, $U^TU_\perp = 0$, and $D_0 \in \mathrm{Sym(p)} = T_IP_p$. The chosen metric of $S^+(p,n)$ is merely the sum of the infinitesimal distance in $\mathrm{Gr(p,n)}$ and in P_p:

$$g_{k(U,R^2)}((\Delta_1, D_1), (\Delta_2, D_2)) \tag{8}$$
$$= \mathrm{Tr}\left(\Delta_1^T\Delta_2\right) + k\ \mathrm{Tr}\left(R^{-1}D_1R^{-2}D_2R^{-1}\right),\ k > 0, \tag{9}$$

generalizing g^{P_n} in a natural way. The space $S^+(p,n) \cong (V_{n,p} \times P_p)/O(p)$ endowed with the metric (8) is a Riemannian manifold, and the metric is invariant to orthogonal transformations, scalings, and pseudo-inversion. The mean of two matrices satisfies the following property:

Proposition 3: In the limit $k \to 0$, the Riemannian mean of A and B with respect to the natural metric (8) is the geometric mean (6), where M is the Ando mean.
This is easy to understand, as when the weight on the cone tends to zero ($k \to 0$), the range of the midpoint matrix between A and B tends to the Grassmannian mean between the ranges of A and B.

On Riemannian manifolds, the geodesic mid-point is the Riemannian mean of two elements. This notion is usually replaced by the Karcher mean for more than two points. For $A_1, ..A_l \in P_n$ viewed as a Riemannian manifold endowed with the natural metric, the Karcher mean is defined to be a minimizer of $X \mapsto \sum_1^l d(X, A_i)^2$. As P_n endowed with the natural metric is a Hadamard manifold, the Karcher mean is uniquely defined and it satisfies all the properties P1-P7. It can be calculated via a simple Newton method on P_n. The proposed mean in this paper is related to the Karcher mean via the following proposition:

Proposition 4: In the limit $k \to 0$, the Karcher mean of $A_1, .., A_l$ with respect to the natural metric (8) is the geometric mean (6), where M is the Ando mean.

The interpretation of the rank-preserving mean (6) as a Riemannian mean on the Riemannian space $(S^+(p,n), g_k)$

is only valid in the limit $k \to 0$. There is no known closed-form for the geodesics of that space and, as a consequence, the Karcher mean can only be computed numerically. The closed-form expression (6) is a good substitute, with the same invariance properties.

V. Geometric means and filtering

A. Filtering a constant rank-p matrix of $\mathrm{S}^+(\mathrm{p}, \mathrm{n})$

In continuous-time, a first-order filter meant to filter a constant noisy signal $y(t)$ writes

$$\tau \frac{d}{dt} x = -x + y$$

In discrete-time, using a semi-implicit numerical scheme, it becomes

$$x_{i+1} = \frac{y_i dt + \tau x_i}{dt + \tau} \tag{10}$$

which is a weighted mean between the measured signal y_i and the filter state x_i. Thus a weighted version of the mean proposed in this paper can be used to do filtering. Indeed if the usual matrix arithmetic mean is used, the rank is not preserved.

For the sake of illustration, we apply this idea to a weighted version of the mean proposed in this paper applied to the filtering of a constant matrix zz^T of $\mathrm{S}^+(1, 2)$, where $z \in \mathbb{R}^2$ is a constant vector. Let $Y(t) = (z + \nu(t))(z + \nu(t))^T \in \mathrm{S}^+(1, 2)$ where $\nu(t) \in \mathbb{R}^2$ is a Gaussian white noise of magnitude 50% of the signal. The filtering algorithm based on a weighted mean corresponding to the metric (8) writes:

$$u_{i+1} = (\cos(\frac{\tau\theta_i + \theta_{Y_i} dt}{dt + \tau}), \sin(\frac{\tau\theta_i + \theta_{Y_i} dt}{dt + \tau})) \tag{11}$$

$$r_{i+1}^2 = \exp(\frac{\tau \log r_i^2 + dt \log s_i^2}{dt + \tau}) \tag{12}$$

where the current estimated matrix is $r_i^2 u_i u_i^T$ with $u_i = (\cos\theta_i, \sin\theta_i)$, and $Y_i = s_i^2 v_i v_i^T$ where $v_i = (\cos\theta_{Y_i}, \sin\theta_{Y_i})$.

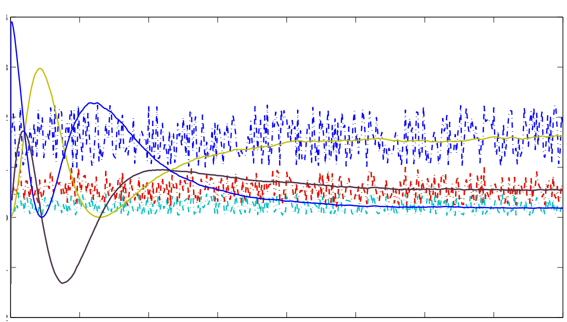

Fig. 1. Filtering on $S^+(1, 2)$: plot of the 3 coefficients of the measured matrix (dashed line) and the filtered matrix (plain line) with a 50% measurement noise, and $\tau = 50dt$. The filter allows to denoise the measured rank-1 symmetric matrix.

B. Diffusion equations on $\mathrm{S}^+(\mathrm{p}, \mathrm{n})$

The mean (6) can be used for filtering images where the pixels are elements of $\mathrm{S}^+(\mathrm{p}, \mathrm{n})$. Indeed one of the most basic filtering algorithm is to convolute the image with a Gaussian kernel. Such a treatment amounts to solve the heat equation on $\mathrm{S}^+(\mathrm{p}, \mathrm{n})$. It is given by

$$\frac{\partial X}{\partial t} = \Delta X$$

Following this article, the problem can be decoupled into a filtering problem on the Grassman manifold of p-dimensional subspaces, and a filtering problem on the set of rank p positive definite matrices. Indeed such an equation can be formulated in discrete time with the help of a Riemannian mean. On a vector space, one can write the Laplacian Δ as the difference between the value at one point, and the (Riemannian) mean value around the point. On a manifold such a difference can be defined as a tangent vector via the exponential map. Such a numerical scheme could maybe yield an approximation of the true Laplacian on the Riemannian manifold (see [13] for a formal definition of the Laplacian), and allow to define a heat equation on the $\mathrm{S}^+(\mathrm{p}, \mathrm{n})$.

Filtering on $S^+(n, n)$ with the metric (8) (which is the GL(n)-invariant metric of the cone $\mathrm{P_n}$) was studied extensively for DTI filtering in [13], [8], [3]. One of the main benefits of this metric is its invariance with respect to scalings which makes it very robust to outliers. Indeed it is easy to see for instance in the $S^+(1, 2)$ case looking at update (12) that an outlier having a very large amplitude will be crushed thanks to the logarithm function. Such a property is inherited by the metric metric (8) for filtering on $\mathrm{S}^+(\mathrm{p}, \mathrm{n})$.

VI. Conclusion

The paper proposes the definition of a rank-preserving geometric mean over the set of fixed-rank positive semidefinite matrices. Our starting point is the framework of Ando [2], which extends arbitrary geometric means from positive scalars to positive definite matrices. The concept can be extended to rank-deficient matrices by a continuity argument but it is well-known that this definition is not rank-preserving. For this reason, we adopt a geometric viewpoint where rank-p positive semidefinite matrices are viewed as "flat" ellipsoids in $\mathbb{R}^n$, i.e. ellipsoids lying in a p-dimensional subspace. The proposed mean computes a mean ellipsoid in the mean p-dimensional subspace. The mean subspace is chosen as a Riemmanian mean on the Grassmann manifold. The mean ellipsoid in the mean subspace is computed from an arbitrary definition of geometric mean between positive definite matrices. Each flat ellipsoid is brought in the mean subspace by a rotation of minimal energy and represented by a (positive definite) matrix coordinate representation in a common basis of the mean subspace. The computation of this matrix representation only involves (thin) SVD calculations. We show that the proposed definition retains all desirable properties of Ando mean, except for the invariance by congruence which is reduced to invariance by scaling and rotation. The proposed mean is interpreted in the context of a Riemannian geometry recently studied over the set of fixed-rank positive semidefinite matrices. Like the Riemannian (Karcher) mean of positive definite matrices, the proposed geometric mean has a closed-form expression for two matrices and can be computed numerically for an arbitratry number of matrices.

APPENDIX

A. Notation

- $\mathrm{P_n}$ is the set of symmetric positive definite $n \times n$ matrices.
- $\mathrm{S^+(p,n)}$ is the set of symmetric positive semidefinite $n \times n$ matrices of rank $p \leq n$. We will only use this notation in the case $p < n$.
- $\mathrm{Gl(n)}$ is the general linear group, that is, the set of invertible $n \times n$ matrices.
- $\mathbb{R}_*^{n\times p}$ is the set of full rank $n \times p$ matrices.
- $\mathrm{V_{n,p}} = \mathrm{O(n)/O(n-p)}$ is the Stiefel manifold; i.e., the set of $n \times p$ matrices with orthonormal columns: $U^T U = I_p$.
- $\mathrm{Gr(p,n)}$ is the Grassman manifold, that is, the set of p-dimensional subspaces of $\mathbb{R}^n$. It can be represented by the equivalence classes $\mathrm{V_{n,p}/O(p)}$.
- $\mathrm{Sym(n)}$ is the vector space of symmetric $n\times n$ matrices.
- $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is the $n \times n$ matrix with the λ_i's on its diagonal. $I = \mathrm{diag}(1, \ldots, 1)$ is the identity matrix.
- $\mathrm{range}(A)$ is the subspace of $\mathbb{R}^n$ spanned by the columns of $A \in \mathbb{R}^{n\times n}$.
- $T_X\mathcal{M}$ is the tangent space to the manifold $\mathcal{M}$ at X.

B. Proofs

Proof of Lemma 1: Suppose $A = z_A z_A^T$ and $B = z_B z_B^T$ are two matrices of $S^+(1,2)$, i.e. $z_A, z_B \in \mathbb{R}^2$. $GL(n)$ congruence invariance means that $M(GAG^T, GBG^T) = GM(A,B)G^T$ for an arbitrary invertible matrix G. If z_A, z_B are not colinear, one can find for each $\epsilon > 0$ an invertible matrix G_ϵ such that $G_\epsilon z_A = z_A$ and $G_\epsilon z_B = \epsilon z_B$. By homogeneity, $M(G_\epsilon AG_\epsilon^T, G_\epsilon BG_\epsilon^T) = \epsilon M(A,B)$, which implies that $G_\epsilon M(A,B)G_\epsilon^T$ tends to zero in the limit $\epsilon \to 0$.

Suppose now $M(A,B) = zz^T$. There exists α, β such that $z = \alpha z_A + \beta z_B$, and we have $G_\epsilon M(A,B)G_\epsilon^T = (\alpha z_A + \epsilon\beta z_B)(\alpha z_A + \epsilon\beta z_B)^T = \alpha^2 z_A z_A^T + O(\epsilon)$. But the zero limit then necessarily implies $\alpha = 0$. Repeating the argument with invertible transformations that only scale z_A instead of z_B leads to the conclusion that $\beta = 0$. Thus if P6 holds $M(A,B) = 0$ and the rank is not preserved.

Proof of Proposition 2: *Proof:* M satisfies P1 and P2 and so does (6). To prove permutation invariance P3 it suffices to note that both Grassman mean and M are unvariant by permutation. To prove P4, suppose $A_i \leq A_i^0$ for each i. Then A_i's and A_i^0 have the same range and they all admit a factorization of the type $WR_i^2W^T$. P4 is then a mere consequence of the monotonicity of M. Using the same arguments, one can prove continuity from above of the mean is a consequence of continuity of M. Finally, P7' can be easily proved noting that for each i the pseudo-inverse writes $A_i^\dagger = U_i R_i^{-2} U_i^T$. Thus the calculation of the mean of the pseudo-inverse yields the inverse Q_i^{-1}'s of the Q_i's and P7' is the consequence of self-duality of M.

Let us prove P6'. As for all $\mu > 0$ and $1 \leq i \leq l$ we have $\mu A_i = U_i(\mu R_i^2)U_i^T$ invariance with respect to scaling is a mere consequence of the invariance of M. Let $O \in \mathrm{O(n)}$. The mean subspace in Grassman of the rotated matrices $\mathrm{range}(OA_iO^T)$'s is the rotated mean subspace of the $\mathrm{range}(A_i)$'s. Moreover U_i W_i are uniquely defined (since $\theta_{i,p-1} < \pi/2$) via SVD (5) for each i, i.e. $U_i^T W_i = D$ with D a diagonal matrix of dimension p. Note that SVD (5) of the rotated matrices yield the rotated matrices OU_i and OW_i as for every i we have $(U_i^T O^T)(OW_i) = U_i^T W_i = D$. Thus the matrices are transformed according to $W_i \mapsto OW_i$ for $0 \leq i \leq l$ and the $Q_i's$ are unchanged. The mean of the rotated matrices is $OW_0\ G(Q_1, ..., Q_l)\ W_0^T O^T$. ■

Proof of Proposition 3: As the object on the paper is not to discuss in great detail the quotient structure of $\mathrm{S^+(p,n)}$ we give a short proof and we suggest the interested reader to look more into detail at Proposition 2 and Theorem 2 of [5]. Let γ be the horizontal lift of a geodesic linking A and B in $\mathrm{S^+(p,n)}$. One can write $\gamma(t) = (U(t), R(t))$ with $A = U(0)R(0)^2U(0)^T$ and $B = U(1)R(1)^2U(1)^T$.

SVD (5) yields representatives of A and B. Consider the horizontal geodesic in the structure space $\mathrm{V_{n,p}} \times \mathrm{P_p}$ joining them $\eta(t) = (V(t), Q(t))$, as in Theorem 2 of [5]. $\mathrm{V_{n,p}} \times \mathrm{P_p}$ is a cartesian product space, and the length of these curves according to the metric (8) write as mere sums. Moreover as they are horizontal curves we have $L_{\mathrm{V_{n,p}}}(\cdot) = L_{\mathrm{Gr(p,n)}}(\cdot)$. Finally

$$\begin{aligned}
L(\gamma) &= L_{\mathrm{Gr(p,n)}}(U) + k\ L_{\mathrm{P_n}}(R) \\
&= L_{\mathrm{Gr(p,n)}}(U) + k\ d_{\mathrm{P_n}}(R_A, R(1)) \\
L(\eta) &= L_{\mathrm{Gr(p,n)}}(V) + k\ L_{\mathrm{P_n}}(Q) \\
&= d_{\mathrm{Gr(p,n)}}(\mathrm{range}(A), \mathrm{range}(B)) + k d_{\mathrm{P_n}}(R_A, Q(1))
\end{aligned}$$

We wrote that $L_{\mathrm{P_n}}(R) = kd_{\mathrm{P_n}}(R_A, R(1))$. Indeed as $L(\gamma)$ is the length of the geodesic γ in $\mathrm{S^+(p,n)}$, $R(t)$ is necessarily a shortest length curve in the cone. Moreover, as the endpoint of both curves in $\mathrm{S^+(p,n)}$ is B, there exists $O \in \mathrm{O(p)}$ such that $U(1)O = V(1)$.

We have $L(\gamma) \leq L(\eta)$ because γ is a geodesic, but also in Grassman $d_{\mathrm{Gr(p,n)}}(\mathrm{range}(A), \mathrm{range}(B)) \leq L_{\mathrm{Gr(p,n)}}(U)$ as U is a curve linking the ranges of A and B in $\mathrm{Gr(p,n)}$. Thus when $k \to 0$, $L_{\mathrm{Gr(p,n)}}(U) - d_{\mathrm{Gr(p,n)}}(\mathrm{range}(A), \mathrm{range}(B))$ tends to 0. As U and V have the same endpoints in Grassman, it implies that $d_{\mathrm{Gr(p,n)}}(U(t), V(t))$ tends uniformly to 0, and so do their horizontal lifts: $d_{\mathrm{V_{n,p}}}(U(t), V(t))$ tends uniformly to 0. In particular we see that as $k \to 0$ the matrix O tends to identity. Thus the geodesics in P_p linking respectively $R_A, Q(1)$ and $R_A, O^T R(1) O$ become infinitely close in the limit. Gathering those last results we see that in particular $d_{\mathrm{S^+(p,n)}}(\eta(1/2), \gamma(1/2))$ tends to zero as $k \to 0$.

Proof of proposition 4: Let us show that in the limit (6) is a minimizer of $X \mapsto \sum_1^l d_{\mathrm{S^+(p,n)}}(X, A_i)^2$. Theorem 2 along with proposition 2 of [5] prove that $|\sum_1^l d_{\mathrm{S^+(p,n)}}(X, A_i)^2 - \sum_1^l d_{\mathrm{Gr(p,n)}}(\mathrm{range}(X), \mathrm{range}(A_i))^2| = O(k)$. Thus the range of the Karcher mean tends to the Grassman mean of the ranges when $k \to 0$.

REFERENCES

[1] T. Ando. Topics on operator inequalities. In *Division of Applied Mathematics, Hokkaido University, Sapporo, Japan*, 1978.

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

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

[4] R. BHATIA, *Positive definite matrices*, Princeton University Press, Princeton, NJ, 2006.

[5] S. Bonnabel and R. Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM J. Matrix Anal. Appl.*, 31(3):1055–1070, 2009.

[6] J. Burbea and C. R. Rao, *Entropy differential metric, distance and divergence measures in probability spaces: A unified approach*, J. Multivariate Anal., 12 (1982), pp. 575–596.

[7] J. Faraut and A. Koranyi, *Analysis on Symmetric Cones*, Oxford University Press, New York, 1994.

[8] P. D. Fletcher and S. Joshi, *Riemannian geometry for the statistical analysis of diffusion tensor data*, Signal Processing, 87 (2007), pp. 250–262.

[9] G.H. Golub and C. Van Loan. *Matrix Computations*. John Hopkins University Press, 1983.

[10] S. Hegalson, *Differential Geometry and Symmetric Spaces*, Pure Appl. Math., Academic Press, New York and London, 1962.

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

[12] M. Zerai and M. Moakher, *The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of diffusion tensor MRI data*, J. Math. Imaging Vision, submitted.

[13] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66:41–66, 2006.

[14] D. Petz and R. Temesi. Means of positive numbers and matrices. *SIAM J. matrix anal. appl.*, 27:712–720, 2005.

[15] W. Pusz and S. L. Woronowicz. Functional calculus for sesquilinear forms and the purification map. *Rep. Mathematical Phys.*, 8:159–170, 1975.

[16] L. T. Skovgaard, *A Riemannian geometry of the multivariate normal model*, Scand. J. Statist., 11 (1984), pp. 211–223.

[17] S.T. Smith. Covariance, subspace, and intrinsic cramer-rao bounds. *IEEE-Transactions on Signal Processing*, 53(5):1610–1629, 2005.