

# A proximal technique for computing the Karcher mean of symmetric positive definite matrices

Ronaldo Malheiros Gregório\* and Paulo Roberto Oliveira†

May 20, 2013

## Abstract

This paper presents a proximal point approach for computing the Riemannian or intrinsic Karcher mean of,  $n \times n$ , *symmetric positive definite (SPD)* matrices. Our method derives from proximal point algorithm with Schur decomposition developed to compute minimum points of convex functions on *SPD* matrices set when it is seen as a Hadamard manifold. The main idea of the original method is preserved. However, here, orthogonal matrices are updated as solutions of optimization subproblems on orthogonal group. Hence, the proximal trajectory is built on solving iteratively Riemannian optimization subproblems alternately on diagonal positive definite matrices set and orthogonal group. No Cholesky factorization is made over variables or datum of the problem. Bunches of numerical experiments, for  $n = 2, \dots, 10$ , and illustrations of the computational behavior of Riemannian gradient descent, proximal point and Richardson-like algorithms are presented at the end.

**Keywords.** Karcher mean, symmetric positive definite matrices, proximal point algorithm, geodesic convexity, orthogonal group.

## 1 Introduction

Denote by  $\mathbb{S}_{++}^n$  the,  $n \times n$ , *SPD* matrices set. Consider the intrinsic Karcher mean (*IKM*) problem of,  $n \times n$ , *SPD* matrices

$$\begin{aligned} & \text{minimize } \frac{1}{2} \sum_{i=1}^m d^2(X^i, X), \\ & X \in \mathbb{S}_{++}^n, \end{aligned}$$

where  $\{X^1, \dots, X^m\} \subset \mathbb{S}_{++}^n$  is a datum set over any hypothetical or real scenery and  $d$ , the Riemannian distance with respect to Riemannian metric defined by Hessian of the standard logarithmic barrier function for  $\mathbb{S}_{++}^n$

$$F(X) = -\ln \det(X). \quad (1)$$

$\mathbb{S}_{++}^n$  is an open Lie subgroup of the,  $n \times n$ , nonsingular matrices set with Riemannian connection and, as consequence of Corollary 5.10 in [21, p. 220], it has nonpositive curvature everywhere. The rule for Riemannian distance between  $A, B \in \mathbb{S}_{++}^n$  resulting from the metric defined by Hessian of (1) is given by

$$d(A, B) = \sqrt{\sum_{l=1}^n \ln^2 \lambda_l(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})}, \quad (2)$$

where  $\lambda_l(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})$  is the  $l^{\text{th}}$  eigenvalue of  $A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$ ,  $l = 1, \dots, n$ .

In the Euclidean sense, means are important tools in statistical process. For instance, let  $\mathcal{P} = \{p_1, \dots, p_m\}$  be a finite subset of any Euclidean vector space. The *Euclidean mean*  $\mu_{\mathcal{P}}$ , defined by

---

\*Universidade Federal Rural do Rio de Janeiro, Departamento de Tecnologias e Linguagens, Av. Gov. Roberto Silveira, s/n, Moquetá, Nova Iguaçu - CEP 26020-740, RJ, Brazil. This research is supported by FAPERJ. E-mail: rgregor@ufrj.br

†Universidade Federal do Rio de Janeiro, Departamento de Eng. de Sistemas e Computação, Caixa Postal 68511, CEP 21945-970, Rio de Janeiro, Brazil. E-mail: poliveir@cos.ufrj.br

$\mu_{\mathcal{P}} = \frac{1}{m} \sum_{i=1}^m p_i$ , the *expected value*  $\sigma_{\mathcal{P}}^2$ , with respect to  $\mu_{\mathcal{P}}$ , given by  $\sigma_{\mathcal{P}}^2 = \sum_{i=1}^m \|p_i - \mu_{\mathcal{P}}\|_2^2$ , where  $\|\cdot\|_2$

is the usual 2-norm for vectors, and the *standard deviation*  $\sigma_{\mathcal{P}} = \sqrt{\sigma_{\mathcal{P}}^2}$  are computed to make an error analysis over  $\mathcal{P}$ .  $\sigma_{\mathcal{P}}^2$  and  $\sigma_{\mathcal{P}}$  represent global error measurements on  $\mathcal{P}$ .

These ideas can be extended for statistical analysis of tensors. The most well-known application belongs in *diffusion tensor (DT)* by *magnetic resonance imaging (MRI)*. In *DT – MRI*, tensors are represented by,  $3 \times 3$ , *SPD* matrices (for further details, see [13]). Statistical analysis is necessary in *DT – MRI* to treat effects of noises over signal intensities. In general, given a set  $\mathcal{T} = \{T_1, \dots, T_m\} \subset$

$\mathbb{S}_{++}^n$  of tensorial data with Euclidean mean  $\mu_{\mathcal{T}} = \frac{1}{m} \sum_{i=1}^m T_i$ , the expected value  $\sigma_{\mathcal{T}}^2$ , defined by  $\sigma_{\mathcal{T}}^2 =$

$\sum_{i=1}^m \text{Tr} \{(T_i - \mu_{\mathcal{T}})^T (T_i - \mu_{\mathcal{T}})\}$ , where  $\text{Tr} \{\cdot\}$  represents the matrix trace, and the standard deviation

$\sigma_{\mathcal{T}} = \sqrt{\sigma_{\mathcal{T}}^2}$  are tools computed to estimate the difference between observed tensors at  $\mathcal{T}$  and the true diffusion tensor. It remarks that  $\mu_{\mathcal{T}}$  can be defined in a continuous differentiable way as the unique solution of the Euclidean Karcher mean (*EKM*) problem

$$\begin{aligned} & \text{minimize } \frac{1}{2} \sum_{i=1}^m \|T_i - \mu\|_F^2 \\ & \mu \in \mathbb{S}_{++}^n, \end{aligned}$$

where  $\|\cdot\|_F$  is the Frobenius norm of matrices defined by  $\|A\|_F = \sqrt{\text{Tr}\{A^T A\}}$ , for any matrix  $A \in \mathbb{R}^{m \times n}$ .

Regarded as *SPD* matrices, tensors can be applied in several research fields. For instance, areas that employ this concept include deformation, image analysis, statistical analysis of diffusion tensor data in medicine, automatic and intelligent control, human detection via classification on the space of *SPD* matrices, pattern recognition, speech emotion classification, modeling of cognitive evolution, analysis of the multifactorial nature of cognitive aging, modeling of functional brain imaging data and design and analysis of wireless cognitive dynamic systems (see [8] for further details).

Nowadays, there are developments with tensors specially in *DT – MRI* working with the Riemannian structure of  $\mathbb{S}_{++}^n$ . For instance, see [5], [9], [14] and [27]. In general, let  $H$  be a Riemannian manifold of nonpositive curvature (or *Hadamard manifold*) and  $\mathcal{H} = \{h_1, \dots, h_m\} \subset H$ . Euclidean Karcher mean is extended to Riemannian case in a single way as the unique solution of the problem

$$\begin{aligned} & \text{minimize } \frac{1}{2} \sum_{i=1}^m d^2(h_i, h) \\ & h \in H, \end{aligned}$$

where  $d$  is the Riemannian distance defined on  $H$ . *IKM* problem of,  $n \times n$ , *SPD* matrices is a particular case of problem above and it differs from *EKM* problem by the choice of the geometrical structure of  $\mathbb{S}_{++}^n$ . Relevant facts listed below are held for *IKM* problem.

$F_1$ .  $\mathbb{S}_{++}^n$  is a Hadamard manifold, with respect to the metric given by Hessian of (1).

$F_2$ . For any  $Y \in \mathbb{S}_{++}^n$ ,  $g(t) = d^2(Y, \gamma(t))$  is a smooth and strictly convex function, for every geodesic segment  $\gamma$  on  $\mathbb{S}_{++}^n$ .

$F_3$ . *IKM* problem has unique solution.

It results from  $F_1$  and Hadamard theorem that  $\mathbb{S}_{++}^n$  is diffeomorphic to  $\mathbb{R}^{\frac{n(n+1)}{2}}$ . In particular, it implies that geodesic segments joining two points in  $\mathbb{S}_{++}^n$  are also unique as in the Euclidean case. The strictly convexity and smoothness in  $F_2$  follow directly from Theorem 6.1.9 (*Semiparallelogram law in*

$\mathbb{S}_{++}^n$ ) in [1, p. 207] and Corollary 6.3.2 in [1, p. 218] or, in an alternative way, as a particular case of Theorem 4.1 in [4, p. 12].  $F_2$  represents the *geodesic strictly convexity property of squared Riemannian distances* in Hadamard manifolds regarding the second argument. The unicity in  $F_3$  results from  $F_2$  and existence, from completeness of  $(\mathbb{S}_{++}^n, d)$  as metric space (Proposition 6.2.2 in [1, p. 210]). Geodesic convexity is discussed with more details in section 3.

The unique solution  $\mu$  for *IKM* problem of,  $n \times n$ , *SPD* matrices is characterized by

$$\sum_{i=1}^m \ln \left( \mu^{-\frac{1}{2}} X^i \mu^{-\frac{1}{2}} \right) = 0. \quad (3)$$

A numerical efficiency measurement for iterative methods to compute  $\mu$  derives from characterization (3). Let  $\Psi$  be an iterative method to solve *IKM* problem and  $\mu_\Psi$ , the approach computed by  $\Psi$ . Here, the *numerical efficiency* of  $\Psi$ , represented by  $EFF_\Psi$ , is defined by  $EFF_\Psi = \left\| \sum_{i=1}^m \ln \left( \mu_\Psi^{-\frac{1}{2}} X^i \mu_\Psi^{-\frac{1}{2}} \right) \right\|_F$ . It results from  $EFF_\Psi$  that, under the same computational conditions (same machine, operational system, software for simulations and stopping criteria, for example), a iterative method  $\Psi_1$  is numerically more efficient than other  $\Psi_2$  if  $EFF_{\Psi_1} < EFF_{\Psi_2}$ .

In relation to methods, different iterative algorithms in nonlinear programming have been extended to Riemannian manifolds. These extensions conglorate classical algorithms as gradient, conjugate gradient and Newton methods (see [9], [18], [23] and [27] for further details). Sophisticated theoretical methods as proximal point algorithms in Hadamard manifolds in [6] and [19] also emphasize this fact. The two late papers are extensions of classical proximal point method introduced in [16] and generalized to monotone operators in [20].

Let  $f : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$  be a geodesic convex function. Given  $X^{(0)} \in \mathbb{S}_{++}^n$  and a sequence of positive real numbers  $\{\beta^{(k)}\}_{k \in \mathbb{N} \cup \{0\}}$ , the theoretical proximal point algorithm, as presented in [6, p. 267], generates a sequence  $\{X^{(k)}\}_{k \in \mathbb{N} \cup \{0\}} \subset \mathbb{S}_{++}^n$  defined by

$$X^{(k+1)} = \operatorname{argmin} \left\{ f(X) + \frac{\beta^{(k)}}{2} d^2(X^{(k)}, X) : X \in \mathbb{S}_{++}^n \right\}, \quad k = 0, 1, \dots, \quad (4)$$

where  $\inf_{X \in \mathbb{S}_{++}^n} \{f(X) + \frac{\beta}{2} d^2(X^{(k)}, X)\}$  extends the Moreau-Yosida regularization to Riemannian case, for any  $\beta > 0$ . Theorem 6.1 in [6, p. 269] holds that  $X^{(k)}$  converges to a minimum point of  $f$  since  $\{\beta^{(k)}\}_{k \in \mathbb{N} \cup \{0\}}$  satisfies  $\sum_{k=0}^{\infty} \frac{1}{\beta^{(k)}} = +\infty$ .

In this paper, the *proximal point algorithm with Schur decomposition* introduced in [11] is fitted to solve the *IKM* problem. Rules for geodesics on diagonal positive definite (*DPD*) matrices set, on orthogonal group and for natural gradient of objective function of *IKM* problem restricted to them are explicitly shown here. Steepest descent algorithm, as presented in [23, p. 117], with Armijo's line search, as in [26, p. 890], is also fitted to solve computationally optimization subproblems in both sets.

Since any information of the objective function of *IKM* problem restricted to orthogonal group is not taken into account in the original version of the algorithm, we emphasize that our method differs from proximal point algorithm with Schur decomposition as far as orthogonal matrices are updated as local solutions of Riemannian optimization subproblems on orthogonal group. No symmetric matrix factorization is required.

In fact, there are a lot of numerical methods to compute eigenvalues and eigenvectors of symmetric matrices, such as power method, inverse iteration, Rayleigh quotient iteration, orthogonal iteration (see [10] for further details). However, as it is still under discussion there, a single *QR* iteration to compute Schur decomposition of symmetric matrices involves  $O(n^3)$  arithmetic flops and its convergence is linear. Other sophisticated methods compute tridiagonal forms for symmetric matrices in  $O(n^2)$  arithmetic flops

and then the shift idea is applied to reduce them to diagonal form. Since our method escapes from Schur factorizations, relatively to proximal point algorithm with Schur decomposition, it is  $O(n^3)$  (or no less than  $O(n^2)$  and no more than  $O(n^3)$ ) arithmetic flops less expensive per internal iteration.

In addition, it is presented a complexity analysis (with respect to a sequence of accuracy parameters  $\epsilon^{(k)} \geq 0$ ,  $k = 0, 1, \dots$ ) for inexact version, numerical experiments in  $\mathbb{S}_{++}^n$  ( $n = 2, 3, \dots, 10$ ) and illustrations of the computational behavior of Riemannian gradient descent method, as presented in [15, p. 2212], proximal point approach, as described here, and Richardson-like algorithm, as introduced in [3, p. 2].

In section 2, we describe some concepts of Riemannian geometry taking into account geometrical structures of  $\mathbb{S}_{++}^n$  and orthogonal group ( $\mathbb{O}_n$ ). In section 3, we discuss the classical proximal point algorithm on Hadamard manifolds, proximal point algorithm with Schur decomposition for geodesic convex problems in  $\mathbb{S}_{++}^n$ , the new orthogonal steps and convergence of internal iterations when applied to *IKM* problem. In section 4, we present an inexact version of the method and its theoretical complexity analysis. In section 5, we compute Riemannian gradient flows of the objective function of *IKM* problem restricted to diagonal positive definite matrices set and  $\mathbb{O}_n$  respectively. Since the aim of this paper is positioning a proximal point method to compute the Riemannian mean of,  $n \times n$ , *SPD* matrices, we improved on bunches of numerical experiments, for  $n = 2, \dots, 10$ , and illustrations that compare the computational behavior of proximal point algorithm with Riemannian gradient descent and Richardson-like algorithms, whose Matlab codes can be found in [2].

## 2 Riemannian aspects of $\mathbb{S}_{++}^n$ and $\mathbb{O}_n$

### 2.1 A brief on the Riemannian structure of $\mathbb{S}_{++}^n$

*SPD* matrices appear in a natural way in several mathematical developments. According to [12], they can be applied in differentiable minimization theory, convexity characterization of differentiable functions, variance-covariance matrices, algebraic and trigonometric moments of nonnegative functions, discretization and difference schemes for numerical solution of differential equations. Some theoretical aspects about  $\mathbb{S}_{++}^n$  follow below.

**Definition 1** Let  $\mathbb{S}^n$  be the real symmetric,  $n \times n$ , matrices set. A matrix  $X \in \mathbb{S}^n$  is said positive definite if  $x^T X x > 0$ , for every  $x \in \mathbb{R}^n$ ,  $x \neq 0$ .

Another concept intrinsically associated with symmetric matrices is presented in the next definition.

**Definition 2** Let  $Q$  be a real,  $n \times n$ , matrix.  $Q$  is said orthogonal if  $Q^T Q = Q Q^T = I$ , where  $I$  is the,  $n \times n$ , identity matrix.

Here, notations  $\mathbb{D}^n$  and  $\mathbb{D}_{++}^n$  are used to denote the diagonal,  $n \times n$ , matrices set and the same set whose elements have strict positive real numbers in its diagonal respectively.

**Theorem 1 (Schur theorem for *SPD* matrices)** Let  $X \in \mathbb{S}^n$ .  $X \in \mathbb{S}_{++}^n$  iff exist matrices  $Q \in \mathbb{O}_n$  and  $\Lambda \in \mathbb{D}_{++}^n$  such that  $X = Q \Lambda Q^T$ .

**Proof.** Theorem 8.1.1 in [10, p. 393] assures existence of  $Q \in \mathbb{O}_n$  and  $\Lambda \in \mathbb{D}^n$  satisfying  $X = Q \Lambda Q^T$ . Since  $\Lambda$  has all eigenvalues of  $X$  in its diagonal, it follows by Theorem 7.2.1 in [12, p. 402] that  $\Lambda \in \mathbb{D}_{++}^n$ . On the contrary, by Sylvester's law of inertia (Theorem 4.5.8 in [12, p. 223]), if exists  $Q \in \mathbb{O}_n$  and  $\Lambda \in \mathbb{D}_{++}^n$  such that  $X = Q \Lambda Q^T$  then  $X \in \mathbb{S}_{++}^n$ . ■

Theorem 1 makes a characterization for  $\mathbb{S}_{++}^n$ . Mathematically,  $\mathbb{S}_{++}^n = \{X \in \mathbb{S}^n | X = Q \Lambda Q^T; Q \in \mathbb{O}_n, \Lambda \in \mathbb{D}_{++}^n\}$ .

**Lemma 1** Let  $X \in \mathbb{S}^n$  and  $f$  be a scalar function defined in  $\lambda_i(X)$ ,  $i = 1, \dots, n$ , where  $\lambda_i(X)$  is the  $i^{\text{th}}$  eigenvalue of  $X$ . Then  $f(X) = Q f(\Lambda) Q^T$ , where  $f(\Lambda) \in \mathbb{D}^n$  satisfies  $[f(\Lambda)]_{ii} = f(\lambda_i(X))$ ,  $i = 1, \dots, n$ .

**Proof.** Against, Theorem 8.1.1 in [10] assures existence of  $Q \in \mathbb{O}_n$  and  $\Lambda \in \mathbb{D}^n$  such that  $X = Q\Lambda Q^T$ , where  $\Lambda_{ii} = \lambda_i(X)$ ,  $i = 1, \dots, n$ . The statement follows as particular case of Corollary 11.1.2 in [10, p. 558].  $\blacksquare$

For instance, Lemma 1 implies that  $e^X = Qe^\Lambda Q^T$  for any  $X \in \mathbb{S}^n$ , where  $Q \in \mathbb{O}_n$  and  $\Lambda \in \mathbb{D}^n$  satisfy  $X = Q\Lambda Q^T$ , and  $\ln(X) = Q\ln(\Lambda)Q^T$ , since  $X \in \mathbb{S}_{++}^n$ . There are no difficulties on computing symmetric and *SPD* matrices functions.

It is well-known from Real Mathematical Analysis that every bounded sequence of positive real numbers has as closure points nonnegative real numbers. This implies that every bounded sequence in  $\mathbb{S}_{++}^n$  has symmetric matrices with nonnegative eigenvalues as closure points. The closure of  $\mathbb{S}_{++}^n$  is called *symmetric positive semidefinite matrices set*.  $\mathbb{S}_{++}^n$  is an open differentiable manifold and its *tangent bundle* is  $T\mathbb{S}_{++}^n = \cup_{X \in \mathbb{S}_{++}^n} T_X \mathbb{S}_{++}^n = \mathbb{S}^n$ , where  $T_X \mathbb{S}_{++}^n$  represents the *tangent space* to  $\mathbb{S}_{++}^n$  at  $X$ .

The real,  $n \times n$ , matrices set is a Hilbert space, with respect to Frobenius scalar product defined by  $\langle A, B \rangle_F = \text{Tr}(A^T B) = \sum_{i,j} a_{ij} b_{ij}$  and  $\mathbb{S}^n$  is a Hilbert subspace of it. Frobenius scalar product restricted to  $\mathbb{S}^n$  is given by  $\langle X, Y \rangle_F = \text{Tr}(XY)$  (with associated norm  $\|X\|_F = \sqrt{\text{Tr}(X^2)}$ ). Since barrier function (1) is strictly convex, with respect to Euclidean metric (metric resulting from Frobenius scalar product), its Hessian induces an inner product on  $T_X \mathbb{S}_{++}^n$ , given by

$$\langle S_1, S_2 \rangle_X = \langle F''(X)S_1, S_2 \rangle_F = \langle X^{-1}S_1X^{-1}, S_2 \rangle_F = \text{Tr} \{ X^{-1}S_1X^{-1}S_2 \}, \quad (5)$$

that depends on  $X \in \mathbb{S}_{++}^n$  in a differentiable way, where  $S_1, S_2 \in T_X \mathbb{S}_{++}^n$  (with associated norm  $\|S\|_X = \sqrt{\text{Tr} \{ (X^{-1}S)^2 \}}$ , for any  $S \in T_X \mathbb{S}_{++}^n$ ).  $\mathbb{S}_{++}^n$  provided itself with the inner product (5) on  $T_X \mathbb{S}_{++}^n$  is a Hadamard manifold.

According to [17, p. 739], the geodesic on  $\mathbb{S}_{++}^n$  satisfying  $\gamma(0) = X$  and  $\gamma'(0) = S$ , for any  $S \in T_X \mathbb{S}_{++}^n$ , is given by

$$\gamma(t) = X^{\frac{1}{2}} e^{tX^{-\frac{1}{2}}SX^{-\frac{1}{2}}} X^{\frac{1}{2}}, \quad t \in \mathbb{R}. \quad (6)$$

In addition, Theorem 6.1.6 in [1, p. 205] establishes the closed rule for geodesic segment  $\gamma$  connecting  $X$  to  $Y$ .  $\gamma$  is explicitly given by

$$\gamma(t) = X^{\frac{1}{2}} \left( X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right)^t X^{\frac{1}{2}}, \quad t \in [0, 1]. \quad (7)$$

The Riemannian distance (2) results from relation  $d(X, Y) = \int_0^1 \|\gamma'(t)\|_{\gamma(t)} dt$ . It follows from (7) that the geodesic segment connecting two diagonal positive definite matrices contains only diagonal positive definite matrices. So, we conclude that  $\mathbb{D}_{++}^n$  is an open subset of  $\mathbb{S}_{++}^n$  that contains the geodesic segment connecting two of their points, for any pair of points in it.

**Definition 3** Let  $f : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$  be a function of class  $C^1$ . The natural (or Riemannian) gradient of  $f$ , denoted by  $\text{grad } f$ , is a map from  $\mathbb{S}_{++}^n$  to  $\mathbb{S}^n$  that satisfies  $\langle \text{grad } f(X), S \rangle_X = \langle \nabla f(X), S \rangle_F$ , for every  $S \in \mathbb{S}^n$ , where  $\nabla f(X)$  is the Euclidean gradient of  $f$  at  $X$ .

Definition 3 is an adaptation of Definition 1.5 in [22, p. 31]. It follows from the inner product (5) that

$$\text{grad } f(X) = X \nabla f(X) X. \quad (8)$$

## 2.2 A brief on the Riemannian structure of $\mathbb{O}_n$

Some statements about  $\mathbb{O}_n$  follow naturally from its own definition. Let  $Q \in \mathbb{O}_n$ . The columns of  $Q$  form an orthonormal basis for  $\mathbb{R}^n$  and  $|\det(Q)| = 1$ .  $\mathbb{O}_n$  is also a Lie subgroup of,  $n \times n$ , nonsingular matrices set with respect to the matrix product. In this section, we cite Riemannian geometrical concepts on  $\mathbb{O}_n$  necessities to state our method.

The Riemannian structure of  $\mathbb{O}_n$  receives especial treatment in [7]. As discussed there,  $\mathbb{O}_n$  is a nonconnected compact manifold that contains a compact connected component. It is known as *special orthogonal group* ( $\mathbb{S}\mathbb{O}_n$ ) and their elements are matrices whose determinant is equal to 1.

For instance, let  $Q(t)$  be a smooth path on  $\mathbb{O}_n$ , with  $Q(0) = Q$ . It follows from Definition 2 that  $\left(\frac{d}{dt}\bigg|_{t=0} Q(t)^T\right) Q(0) + Q(0)^T \left(\frac{d}{dt}\bigg|_{t=0} Q(t)\right) = (Q'(0))^T Q + Q^T Q'(0) = 0$ . This implies that  $T_Q \mathbb{O}_n = \{X \in \mathbb{R}^{n \times n} | X^T Q + Q^T X = 0\}$ . Moreover, the inner product in  $T_Q \mathbb{O}_n$  coincides with Frobenius scalar product for,  $n \times n$ , matrices. The rule to the geodesic  $\xi$  in  $\mathbb{O}_n$ , satisfying  $\xi(0) = Q$  and  $\xi'(0) = V$ , for any  $V \in T_Q \mathbb{O}_n$ , is given by

$$\xi(t) = \exp(tVQ^T)Q, \quad (9)$$

where  $\exp(tX) = \sum_{k=0}^{\infty} \frac{(tX)^k}{k!}$ . The rule for natural gradient of a function  $f : \mathbb{O}_n \rightarrow \mathbb{R}$ , of class  $C^1$ , at  $T_Q \mathbb{O}_n$  is given by

$$\text{grad } f(Q) = \frac{1}{2}(\nabla f(Q) - Q\nabla f(Q)^T Q), \quad (10)$$

where  $\nabla f(Q)$  is the Euclidean gradient of  $f$  at  $Q$ .

## 3 IKM problem and proximal point algorithm

### 3.1 Convexity and proximal regularization

In this section, proximal results are discussed with respect to Riemannian convexity in  $\mathbb{S}_{++}^n$ . All the statements about Euclidean convexity can be extended in a Riemannian sense for  $\mathbb{S}_{++}^n$ , since it represents an example of Hadamard manifolds. We recommend [4] for further explanations about geodesic convexity in Hadamard manifolds. Properties of geodesic convex programs in Riemannian manifolds as, for example, characterization of minimum points to smooth and geodesic convex function and unicity of interior minimum point for smooth and strictly convex function (if it exists) can be found in [25]. By now, we are interested in synthesizing extensions of Euclidean convexity properties used by our method.

**Definition 4 (Geodesic convex sets and functions)** *Let  $C \subseteq \mathbb{S}_{++}^n$ .  $C$  is said geodesic convex if for every  $X, Y \in C$ , the geodesic segment  $\gamma$  connecting  $X$  to  $Y$  belongs on  $C$ . In addition,  $g : C \rightarrow \mathbb{R}$  is said geodesic (strictly) convex on  $C$  if  $g(\gamma(t))$  is a (strictly) convex function, for every geodesic segment  $\gamma$  in  $C$ .*

**Lemma 2** *Let  $Y \in \mathbb{S}_{++}^n$ . Define  $f_Y : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$  by  $f_Y(X) = \frac{1}{2}d^2(Y, X)$ .  $f_Y$  is a smooth and geodesic strictly convex function and its Riemannian gradient is given by*

$$\text{grad } f_Y(X) = -X^{\frac{1}{2}} \ln \left( X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right) X^{\frac{1}{2}}. \quad (11)$$

**Proof.** Since  $\mathbb{S}_{++}^n$  is a Hadamard manifold, geodesic strictly convexity and smoothness of  $f_Y$  is held. Let  $\gamma_X$  be the geodesic segment on  $\mathbb{S}_{++}^n$  connecting  $Y$  to  $X$  ( $\gamma_X(0) = Y$ ,  $\gamma_X(1) = X$  and  $\gamma_X(t) \in \mathbb{S}_{++}^n$ ,  $\forall t \in (0, 1)$ ) and  $\tilde{\gamma}_X$ , its reparametrization by arc length. From Proposition 4.8 in [22, p. 108] we have that the gradient of  $d_Y : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$ , defined by  $d_Y(X) = d(Y, X)$ , is given by  $\text{grad } d_Y(X) = \tilde{\gamma}'_X(d(Y, X)) = \frac{1}{d(Y, X)} \gamma'_X(1)$ . Let  $W \in \mathbb{S}^n$ . Put  $r(t) = f_Y(\eta(t)) = \frac{1}{2}d^2(Y, \eta(t))$ , for any smooth path  $\eta$  in  $\mathbb{S}_{++}^n$  tangent to  $W$  at  $\eta(0) = X$ . We have that  $\langle \text{grad } f_Y(X), W \rangle_X = \frac{d}{dt} \bigg|_{t=0} r(t) = \frac{1}{2} [2d(Y, \eta(t)) \langle \text{grad } d_Y(\eta(t)), \eta'(t) \rangle_{\eta(t)}] \bigg|_{t=0} = \langle \gamma'_X(1), W \rangle_X$ . Hence,  $\text{grad } f_Y(X) = \gamma'_X(1)$ . To finish, define  $\tilde{\gamma}_X(t) = \gamma_X(1 - t)$ . It is easily seen that  $\tilde{\gamma}_X$  is the geodesic segment from  $X$  to  $Y$  and

$\tilde{\gamma}'_X(t) = -\gamma'_X(1-t)$ . This implies that  $\gamma'_X(1) = -\exp_X^{-1}Y$ , where  $\exp_X^{-1}Y$  is the velocity of  $\tilde{\gamma}_X$  at  $X$  ( $\exp_X^{-1}Y = \tilde{\gamma}'_X(0)$ ). Equality (11) comes from differentiation of (7), with respect to  $t$ , at  $t = 0$ . ■

It is well-known that differentiable functions in Euclidean spaces are convex iff they satisfy the *gradient's inequality*. This is still guaranteed for geodesic convex function on Hadamard manifolds. In particular, the following statement holds it for  $\mathbb{S}_{++}^n$ .

**Lemma 3** *Let  $\mathcal{S} \subseteq \mathbb{S}_{++}^n$  be geodesic convex and  $g : \mathcal{S} \rightarrow \mathbb{R}$ , a function of class  $C^1$ .  $g$  is geodesic convex on  $\mathcal{S}$  iff  $g(X) \geq g(Y) + \langle \text{grad } g(Y), \exp_Y^{-1}X \rangle_Y$ , for every  $X, Y \in \mathcal{S}$ . In addition,  $g$  is geodesic strictly convex iff the inequality above is strict, for every  $X \neq Y$ .*

**Proof.** It follows as a particular case of Theorem 5.1 in [25, p. 78]. ■

**Definition 5** *Let  $Y \in \mathbb{S}_{++}^n$ .  $g : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$  is said 1-coercive at  $Y$  if  $\lim_{d(Y,X) \rightarrow +\infty} \frac{g(X)}{d(Y,X)} = +\infty$ .*

It remarks that 1-coercivity, under the continuity assumption for  $g$ , holds existence of minimum points. In fact, definition above implies that, for every sequence  $\{Y^{(k)}\}_{k \in \mathbb{N}}$ , such that  $d(Y, Y^{(k)}) \rightarrow +\infty$ ,  $g(Y^{(k)}) \geq d(Y, Y^{(k)})$  for  $k$  sufficiently great. Therefore,  $g$  attains its minimum in a closed geodesic ball around  $Y$  ( $B_\epsilon(Y) = \{X \in \mathbb{S}_{++}^n | d(Y, X) \leq \epsilon\}$ , for any  $\epsilon > 0$ ). Since closed geodesic balls are compact, affirmation at the beginning of the paragraph follows.

**Lemma 4** *Let  $g : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$  be a smooth and geodesic convex function and  $Y \in \mathbb{S}_{++}^n$ . Then  $(g + \beta f_Y)$  is 1-coercive at  $Y$ .*

**Proof.** Since  $g$  is differentiable,  $\partial g(Y) = \{\text{grad } g(Y)\}$ , where  $\partial g(Y)$  is the subdifferential of  $g$  at  $Y$ . On the other hand, smoothness and geodesic convexity of  $g$  holds Lemma 3. At this point the argumentation follows the same steps of Lemma 4.1 proof's in [6, p. 266]. ■

**Corollary 1** *Let  $f : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$  be the function given by  $f(X) = \sum_{i=1}^m f_{X^i}(X)$ . For each  $Y \in \mathbb{S}_{++}^n$  and  $\beta > 0$ , define  $(f + \beta f_Y) : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$  by  $(f + \beta f_Y)(X) = f(X) + \beta f_Y(X)$ . Then both,  $f$  and  $(f + \beta f_Y)$  are smooth and geodesic strictly convex functions, with the second 1-coercive at  $Y$ , and their Riemannian gradients are given by  $\text{grad } f(X) = -X^{\frac{1}{2}} \sum_{i=1}^m \ln \left( X^{-\frac{1}{2}} X^i X^{-\frac{1}{2}} \right) X^{\frac{1}{2}}$  and  $\text{grad } (f + \beta f_Y)(X) = -X^{\frac{1}{2}} \sum_{i=1}^m \ln \left( X^{-\frac{1}{2}} X^i X^{-\frac{1}{2}} \right) X^{\frac{1}{2}} - \beta X^{\frac{1}{2}} \ln \left( X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right) X^{\frac{1}{2}}$  respectively.*

**Proof.** It follows from Lemma 2 that  $f$  and  $(f + \beta f_Y)$  are sums of smooth and geodesic strictly convex functions. Hence,  $f$  and  $(f + \beta f_Y)$  are also smooth and geodesic strictly convex functions. In addition, it follows, from Lemma 4, that  $(f + \beta f_Y)$  is 1-coercive at  $Y$ .  $\text{grad } f(X)$  and  $\text{grad } (f + \beta f_Y)(X)$  follow from (11). ■

**Proposition 1** *Let  $\{X^{(k)}\}_{k \in \mathbb{N} \cup \{0\}}$  be the sequence generated by (4).  $X^{(k+1)}$  is well-defined and characterized by*

$$\sum_{i=1}^m \ln \left( X^{(k+1)-\frac{1}{2}} X^i X^{(k+1)-\frac{1}{2}} \right) = -\beta^{(k)} \ln \left( X^{(k+1)-\frac{1}{2}} X^{(k)} X^{(k+1)-\frac{1}{2}} \right). \quad (12)$$

**Proof.** Since  $X^{(k+1)}$  is defined as solution of iteration (4),  $\text{grad } (f + \beta^{(k)} f_{X^{(k)}})(X^{(k+1)}) = 0$  (see Theorem 7.5 in [25, p. 91]). The statement follows from the expression of  $\text{grad } (f + \beta f_Y)$  in Corollary 1 and nonsingularity of  $X^{(k+1)}$ . ■

**Lemma 5** *There is only one solution  $\mu$  for IKM problem characterized by (3).*

**Proof.** See [1, p. 217]. ■

### 3.2 *SDPProx* algorithm and new orthogonal steps

Now, the proximal point algorithm in [11, p. 473], named *SDPProx* algorithm, can be fitted to compute the intrinsic Karcher mean of,  $n \times n$ , *SPD* matrices, since it is defined as the unique solution of a geodesic strictly convex optimization problem on  $\mathbb{S}_{++}^n$ . In fact, that algorithm was developed to compute minimum points of geodesic convex functions defined on  $\mathbb{S}_{++}^n$ .

Its convergence is held under no assumption of differentiability there because the subdifferential theory for geodesic convex function on Hadamard manifolds is already well positioned (see [6] or [25] for further details). It remarks that *SDPProx* algorithm is a variant of the theoretical proximal point method in Hadamard manifolds, where  $X^{(k+1)}$ , defined at iteration (4), is iteratively computed in two recursive phases. For instance, given  $Q_{(0)}^{(k)} \in \mathbb{O}_n$  and  $\Lambda_{(0)}^{(k)} \in \mathbb{D}_{++}^n$ , for any  $k$  ( $k = 0, 1, \dots$ ), the internal loop of *SDPProx* algorithm generates sequences  $\{Q_{(j)}^{(k)}\}_{j \in \mathbb{N} \cup \{0\}} \subset \mathbb{O}_n$  and  $\{\Lambda_{(j)}^{(k)}\}_{j \in \mathbb{N} \cup \{0\}} \subset \mathbb{D}_{++}^n$  by the following steps

$$S1. \bar{\Lambda}_{(j+1)}^{(k)} = \operatorname{argmin} \left\{ \phi_{(j)}^{(k)}(\Lambda) + \beta^{(k)} \rho_I(\Lambda) : \Lambda \in \mathbb{D}_{++}^n \right\},$$

$$S2. \left( \Lambda_{(j+1)}^{(k)}, Q_{(j+1)}^{(k)} \right) \leftarrow \mathbf{Schur} \left( Q_{(j)}^{(k)} \bar{\Lambda}_{(j+1)}^{(k)} Q_{(j)}^{(k)T} \right), j = 0, 1, \dots,$$

where  $\phi_{(j)}^{(k)}(\Lambda) = f \left( X^{(k)\frac{1}{2}} Q_{(j)}^{(k)} \Lambda Q_{(j)}^{(k)T} X^{(k)\frac{1}{2}} \right)$ ,  $\rho_I(\Lambda) = f_{X^{(k)}} \left( X^{(k)\frac{1}{2}} Q_{(j)}^{(k)} \Lambda Q_{(j)}^{(k)T} X^{(k)\frac{1}{2}} \right)$  and, for any  $X \in \mathbb{S}_{++}^n$ , *Schur*( $X$ ) returns matrices  $\Lambda \in \mathbb{D}_{++}^n$  and  $Q \in \mathbb{O}_n$  as in Theorem 1. Proposition 2 in [11, p. 475] holds the convergence of the sequence  $\{Y_{(j)}^{(k)}\}_{j \in \mathbb{N} \cup \{0\}} \subset \mathbb{S}_{++}^n$ , defined by the iteration

$$Y_{(j)}^{(k)} = X^{(k)\frac{1}{2}} Q_{(j)}^{(k)} \Lambda_{(j)}^{(k)} Q_{(j)}^{(k)T} X^{(k)\frac{1}{2}}, \quad (13)$$

to  $X^{(k+1)}$ , since  $\bar{\Lambda}_{(j+1)}^{(k)} \neq \Lambda_{(j)}^{(k)}$ , for every  $j = 0, 1, \dots$ .

It remarks that *SDPProx* algorithm replaces iteration 4 by a sequence of problems defined on diagonal positive definite matrices set followed by Schur decompositions to update both, diagonal positive definite and orthogonal iterated. It represents a reduction of  $n(n-1)/2$  variables per internal iterations compared to iteration (4). Unfortunately, no practical way to update  $Q_{(j)}^{(k)}$  is discussed in [11] if  $\bar{\Lambda}_{(j+1)}^{(k)} = \Lambda_{(j)}^{(k)}$ . In addition, the updating from  $Q_{(j)}^{(k)}$  to  $Q_{(j+1)}^{(k)}$  in the *SDPProx* algorithm does not take into account informations of  $\phi_{(j)}^{(k)}$ , with respect to the Riemannian structure of  $\mathbb{O}_n$ . Moreover, it is easily seen that  $\rho_I$  does not depend on  $Q$ . Next, we will improve a new updating for orthogonal iterated.

Firstly, it observes that the rings  $(\mathbb{D}^n, +, \cdot)$  and  $(\mathbb{R}^n, \oplus, \odot)$  (in particular, the fields  $(\mathbb{D}_{++}^n, +, \cdot)$  and  $(\mathbb{R}_{++}^n, \oplus, \odot)$ ) have the same algebraic structure in sense of the ring isomorphism  $diag : (\mathbb{D}^n, +, \cdot) \rightarrow (\mathbb{R}^n, \oplus, \odot)$  defined by  $diag(\Lambda) = (\Lambda_{11}, \dots, \Lambda_{nn})^T$ , where  $\mathbb{R}^n$  represents the Euclidean vector space of dimension  $n$ ,  $+$  and  $\cdot$ , the standard addition and product of matrices and  $\oplus$  and  $\odot$ , the standard addition and Hadamard product of vectors given by  $x \odot y = (x_1 y_1, \dots, x_n y_n)^T$ , for any  $x, y \in \mathbb{R}^n$  (in particular,  $\mathbb{R}_{++}^n$  represents the positive orthant of  $\mathbb{R}^n$ ).

**Remark 1** For simplicity, the capital letter  $\Lambda$  is employed to represent diagonal matrices and the corresponding small letter  $\lambda$ , for vectors whose components are diagonal elements of  $\Lambda$  in the order that they appear in its diagonal.

Let  $\phi^{(k)} : \mathbb{R}_{++}^n \times \mathbb{O}_n \rightarrow \mathbb{R}$  be the function defined by

$$\phi^{(k)}(\lambda, Q) = f \left( X^{(k)\frac{1}{2}} Q \Lambda Q^T X^{(k)\frac{1}{2}} \right) = \frac{1}{2} \sum_{i=1}^m d^2 \left( X^i, X^{(k)\frac{1}{2}} Q \Lambda Q^T X^{(k)\frac{1}{2}} \right). \quad (14)$$

Based on these facts,  $Q_{(j+1)}^{(k)}$  is defined in an alternative way as a minimizer of  $\phi^{(k)}(\lambda_{(j+1)}^{(k)}, \cdot)$ , since the minimizer set of  $\phi^{(k)}(\lambda_{(j+1)}^{(k)}, \cdot)$  on  $\mathbb{O}_n$  is nonempty (notice that  $\phi^{(k)}(\lambda_{(j+1)}^{(k)}, \cdot)$  is continuous on  $\mathbb{O}_n$



and itself is compact). The alternative version for *SDPProx* algorithm, originally called by *proximal point algorithm* here, rewrites *S1* as

$$\lambda_{(j+1)}^{(k)} = \operatorname{argmin} \left\{ \phi^{(k)}(\lambda, Q_{(j)}^{(k)}) + \beta^{(k)} \rho_I(\lambda) : \lambda \in \mathbb{R}_{++}^n \right\}, \quad (15)$$

and it replaces *S2* by

$$Q_{(j+1)}^{(k)} \in \operatorname{argmin} \left\{ \phi^{(k)}(\lambda_{(j+1)}^{(k)}, Q) : Q \in \mathbb{O}_n \right\}, j = 0, 1, \dots. \quad (16)$$

### 3.3 Internal convergence analysis

**Proposition 2** *Let  $\{\Lambda_{(j)}^{(k)}\}_{j \in \mathbb{N} \cup \{0\}}$  and  $\{Q_{(j)}^{(k)}\}_{j \in \mathbb{N} \cup \{0\}}$  be sequences generated by (15) and (16) respectively. Set  $Y_{(j)}^{(k)}$  as in (13) and  $\bar{Y}_{(j+1)}^{(k)} = X^{(k)\frac{1}{2}} Q_{(j)}^{(k)} \Lambda_{(j+1)}^{(k)} Q_{(j)}^{(k)T} X^{(k)\frac{1}{2}}$ . Then,  $(f + f_{X^{(k)}})(Y_{(j)}^{(k)}) \geq (f + f_{X^{(k)}})(Y_{(j+1)}^{(k)}) + \frac{1}{2} d^2(Y_{(j)}^{(k)}, \bar{Y}_{(j+1)}^{(k)})$ , for every  $j \in \mathbb{N} \cup \{0\}$ .*

**Proof.** In fact, it follows by Lemma 4 in [11, p. 474] that

$$d^2(\Lambda_{(j)}^{(k)}, \Lambda_{(j+1)}^{(k)}) + d^2(\Lambda_{(j+1)}^{(k)}, I) - \frac{2}{\beta^{(k)}} \left( \phi^{(k)}(\Lambda_{(j)}^{(k)}, Q_{(j)}^{(k)}) - \phi^{(k)}(\Lambda_{(j+1)}^{(k)}, Q_{(j)}^{(k)}) \right) \leq d^2(\Lambda_{(j)}^{(k)}, I).$$

Multiplying inequality above by  $\frac{\beta^{(k)}}{2}$ , we have that

$$\phi^{(k)}(\Lambda_{(j+1)}^{(k)}, Q_{(j)}^{(k)}) + \frac{\beta^{(k)}}{2} d^2(\Lambda_{(j+1)}^{(k)}, I) + \frac{\beta^{(k)}}{2} d^2(\Lambda_{(j)}^{(k)}, \Lambda_{(j+1)}^{(k)}) \leq \phi^{(k)}(\Lambda_{(j)}^{(k)}, Q_{(j)}^{(k)}) + \frac{\beta^{(k)}}{2} d^2(\Lambda_{(j)}^{(k)}, I).$$

Since the Riemannian distance (2) is invariant under isometries (see Lemma 6.1.1 in [1, p. 201]), applying the isometry  $\Gamma_{Q_{(j)}^{(k)T} X^{(k)\frac{1}{2}}} : \mathbb{S}_{++}^n \rightarrow \mathbb{S}_{++}^n$ , given by  $\Gamma_{Q_{(j)}^{(k)T} X^{(k)\frac{1}{2}}}(Y) = X^{(k)\frac{1}{2}} Q_{(j)}^{(k)} Y Q_{(j)}^{(k)T} X^{(k)\frac{1}{2}}$ ,

and definition of  $\phi^{(k)}$  in the last inequality, we have that

$$f(Y_{(j)}^{(k)}) + f_{X^{(k)}}(Y_{(j)}^{(k)}) \geq f(\bar{Y}_{(j+1)}^{(k)}) + f_{X^{(k)}}(\bar{Y}_{(j+1)}^{(k)}) + \frac{1}{2} d^2(Y_{(j)}^{(k)}, \bar{Y}_{(j+1)}^{(k)}). \quad (17)$$

On the other hand,  $f(\bar{Y}_{(j+1)}^{(k)}) = f\left(X^{(k)\frac{1}{2}} Q_{(j)}^{(k)} \Lambda_{(j+1)}^{(k)} Q_{(j)}^{(k)T} X^{(k)\frac{1}{2}}\right) = \phi\left(\lambda_{(j+1)}^{(k)}, Q_{(j)}^{(k)}\right)$   
 $\geq \phi\left(\lambda_{(j+1)}^{(k)}, Q_{(j+1)}^{(k)}\right) = f\left(X^{(k)\frac{1}{2}} Q_{(j+1)}^{(k)} \Lambda_{(j+1)}^{(k)} Q_{(j+1)}^{(k)T} X^{(k)\frac{1}{2}}\right) = f(Y_{(j+1)}^{(k)})$  and  $f_{X^{(k)}}(\bar{Y}_{(j+1)}^{(k)})$   
 $= f_{X^{(k)}}\left(X^{(k)\frac{1}{2}} Q_{(j)}^{(k)} \Lambda_{(j+1)}^{(k)} Q_{(j)}^{(k)T} X^{(k)\frac{1}{2}}\right) = \frac{1}{2} d^2\left(X^{(k)}, X^{(k)\frac{1}{2}} Q_{(j)}^{(k)} \Lambda_{(j+1)}^{(k)} Q_{(j)}^{(k)T} X^{(k)\frac{1}{2}}\right)$   
 $= \frac{1}{2} \sum_{l=1}^n \ln^2 \theta_l\left(Q_{(j)}^{(k)} \Lambda_{(j+1)}^{(k)} Q_{(j)}^{(k)T}\right) = \frac{1}{2} \sum_{l=1}^n \ln^2 \theta_l\left(\Lambda_{(j+1)}^{(k)}\right) = \frac{1}{2} \sum_{l=1}^n \ln^2 \theta_l\left(Q_{(j+1)}^{(k)} \Lambda_{(j+1)}^{(k)} Q_{(j+1)}^{(k)T}\right)$   
 $= \frac{1}{2} d^2\left(X^{(k)}, X^{(k)\frac{1}{2}} Q_{(j+1)}^{(k)} \Lambda_{(j+1)}^{(k)} Q_{(j+1)}^{(k)T} X^{(k)\frac{1}{2}}\right) = f_{X^{(k)}}\left(X^{(k)\frac{1}{2}} Q_{(j+1)}^{(k)} \Lambda_{(j+1)}^{(k)} Q_{(j+1)}^{(k)T} X^{(k)\frac{1}{2}}\right)$   
 $= f_{X^{(k)}}(Y_{(j+1)}^{(k)}).$  So, we conclude that

$$f(\bar{Y}_{(j+1)}^{(k)}) \geq f(Y_{(j+1)}^{(k)}) \quad \text{and} \quad f_{X^{(k)}}(\bar{Y}_{(j+1)}^{(k)}) = f_{X^{(k)}}(Y_{(j+1)}^{(k)}). \quad (18)$$

Replacing both relations at (18) in the inequality (17), the statement follows.  $\blacksquare$

**Corollary 2** *If  $\Lambda_{(j+1)}^{(k)} \neq \Lambda_{(j)}^{(k)}$  then*

$$(f + f_{X^{(k)}})(Y_{(j)}^{(k)}) > (f + f_{X^{(k)}})(Y_{(j+1)}^{(k)}). \quad (19)$$

**Proof.** Since  $\Lambda_{(j+1)}^{(k)} \neq \Lambda_{(j)}^{(k)}$ ,  $0 < d\left(\Lambda_{(j+1)}^{(k)}, \Lambda_{(j)}^{(k)}\right) = d\left(Y_{(j)}^{(k)}, \bar{Y}_{(j+1)}^{(k)}\right)$ . So, inequality (19) results from Proposition 2.  $\blacksquare$

**Corollary 3** *Suppose that the hypothesis of Corollary 2 breaks. If  $\phi^{(k)}\left(\lambda_{(j)}^{(k)}, Q_{(j+1)}^{(k)}\right) < \phi^{(k)}\left(\lambda_{(j)}^{(k)}, Q_{(j)}^{(k)}\right)$  then inequality (19) is still held.*

**Proof.** Since the hypothesis of Corollary 2 is not true, the equality  $\Lambda_{j+1}^k = \Lambda_j^k$  implies that  $\bar{Y}_{(j+1)}^{(k)} = Y_{(j)}^{(k)}$  and, consequently,  $d^2\left(Y_{(j)}^{(k)}, \bar{Y}_{(j+1)}^{(k)}\right) = 0$ . However,  $\phi^{(k)}\left(\lambda_{(j+1)}^{(k)}, Q_{(j+1)}^{(k)}\right) = \phi^{(k)}\left(\lambda_{(j)}^{(k)}, Q_{(j+1)}^{(k)}\right) < \phi^{(k)}\left(\lambda_{(j)}^{(k)}, Q_{(j)}^{(k)}\right)$ . This implies that  $f\left(Y_{(j+1)}^{(k)}\right) < f\left(Y_{(j)}^{(k)}\right)$ . Taking into account the latest inequality and the equality in (18), inequality (19) still follows.  $\blacksquare$

**Proposition 3** *Let  $\{\Lambda_{(j)}^{(k)}\}_{j \in \mathbb{N} \cup \{0\}}$  and  $\{Q_{(j)}^{(k)}\}_{j \in \mathbb{N} \cup \{0\}}$  be the sequences generated by (15) and (16) respectively. Suppose that the following two statements are held*

$$(i) \Lambda_{(j+1)}^{(k)} = \Lambda_{(j)}^{(k)},$$

$$(ii) \phi^{(k)}\left(\lambda_{(j)}^{(k)}, Q_{(j)}^{(k)}\right) = \phi^{(k)}\left(\lambda_{(j)}^{(k)}, Q_{(j+1)}^{(k)}\right) \text{ for every minimizer } Q_{(j+1)}^{(k)} \text{ of } \phi^{(k)}\left(\lambda_{(j)}^{(k)}, \cdot\right).$$

Then  $Y_j^{(k)} = X^{(k+1)}$ .

**Proof.** By contradiction, suppose that  $Y_{(j)}^{(k)} \neq X^{(k+1)}$ . Since  $(f + \beta^{(k)}f_{X^{(k)}})$  is geodesic strictly convex, for every  $\delta > 0$ , the normal ball around  $Y_{(j)}^{(k)}$ , given by  $B_\delta\left(Y_{(j)}^{(k)}\right) = \left\{Y \in \mathbb{S}_{++}^n : d\left(Y_{(j)}^{(k)}, Y\right) < \delta\right\}$ , contains at least one point  $Y_\delta$  such that

$$\left(f + \beta^{(k)}f_{X^{(k)}}\right)\left(Y_\delta\right) < \left(f + \beta^{(k)}f_{X^{(k)}}\right)\left(Y_{(j)}^{(k)}\right) \quad (20)$$

(if this is not true,  $Y_{(j)}^{(k)}$  would be a local minimum of  $(f + \beta^{(k)}f_{X^{(k)}})$  and, as consequence of its geodesic strict convexity,  $Y_{(j)}^{(k)} = X^{(k+1)}$ ). By Theorem 1 and taking into account that  $\Gamma_{X^{(k)}\frac{1}{2}}(Y) = X^{(k)\frac{1}{2}}YX^{(k)\frac{1}{2}}$  is an isometry on  $\mathbb{S}_{++}^n$ , there are matrices  $Q_\delta \in \mathbb{O}_n$  and  $\Lambda_\delta \in \mathbb{D}_{++}^n$  such that  $Y_\delta = X^{(k)\frac{1}{2}}Q_\delta\Lambda_\delta Q_\delta^T X^{(k)\frac{1}{2}}$ . Now, we can divide the proof in two cases. Firstly, admit that  $Q_\delta = Q_{(j)}^{(k)}$ . It follows from inequality (20) that  $\Lambda_\delta \neq \Lambda_{(j)}^{(k)}$  and  $\phi^{(k)}\left(\lambda_\delta, Q_{(j)}^{(k)}\right) + \beta^{(k)}\rho_I\left(\lambda_\delta\right) = \phi^{(k)}\left(\lambda_\delta, Q_\delta\right) + \beta^{(k)}\rho_I\left(\lambda_\delta\right) = (f + \beta^{(k)}f_{X^{(k)}})\left(Y_\delta\right) < (f + \beta^{(k)}f_{X^{(k)}})\left(Y_{(j)}^{(k)}\right) = \phi^{(k)}\left(\lambda_{(j)}^{(k)}, Q_{(j)}^{(k)}\right) + \beta^{(k)}\rho_I\left(\lambda_{(j)}^{(k)}\right)$ . It contradicts (i). Now, suppose that  $Q_\delta \neq Q_{(j)}^{(k)}$ . In addition, admit that  $\Lambda_\delta \neq \Lambda_{(j)}^{(k)}$ . Let  $\gamma$  be the geodesic segment from  $Y_{(j)}^{(k)}$  to  $Y_\delta$ . We have that  $(f + \beta^{(k)}f_{X^{(k)}})(\gamma(t)) < (1-t)(f + \beta^{(k)}f_{X^{(k)}})(Y_{(j)}^{(k)}) + t(f + \beta^{(k)}f_{X^{(k)}})(Y_\delta) < (f + \beta^{(k)}f_{X^{(k)}})(Y_{(j)}^{(k)})$ , for every  $t \in (0, 1)$ . In particular,  $(f + \beta^{(k)}f_{X^{(k)}})(\gamma(\frac{1}{2})) < (f + \beta^{(k)}f_{X^{(k)}})(Y_{(j)}^{(k)})$ . On the other side, against by Theorem 1 and the isometry  $\Gamma_{X^{(k)}\frac{1}{2}}$ , for every  $t \in (0, 1)$  there are  $Q(t) \in \mathbb{O}_n$  and  $\Lambda(t) \in \mathbb{D}_{++}^n$ , such that  $\gamma(t) = X^{(k)\frac{1}{2}}Q(t)\Lambda(t)Q^T(t)X^{(k)\frac{1}{2}}$ . Setting  $Q(0) = Q_{(j)}^{(k)}$ ,  $\Lambda(0) = \Lambda_{(j)}^{(k)}$ ,  $Q(1) = Q_\delta$  and  $\Lambda(1) = \Lambda_\delta$ , we have that  $Q(t)$  and  $\Lambda(t)$  are paths in  $\mathbb{O}_n$  and  $\mathbb{D}_{++}^n$  from  $Q_{(j)}^{(k)}$  to  $Q_\delta$  and from  $\Lambda_{(j)}^{(k)}$  to  $\Lambda_\delta$ , respectively. Set  $\bar{Y}_\delta = X^{(k)\frac{1}{2}}Q_\delta\Lambda_{(j)}^{(k)}Q_\delta^T X^{(k)\frac{1}{2}}$ . By Semiparallelogram law,

$$d^2\left(\gamma\left(\frac{1}{2}\right), \bar{Y}_\delta\right) \leq \frac{1}{2}d^2\left(Y_{(j)}^{(k)}, \bar{Y}_\delta\right) + \frac{1}{2}d^2\left(Y_\delta, \bar{Y}_\delta\right) - \frac{1}{4}d^2\left(Y_{(j)}^{(k)}, Y_\delta\right) < \frac{1}{2}d^2\left(Y_{(j)}^{(k)}, \bar{Y}_\delta\right) + \frac{1}{2}d^2\left(Y_\delta, \bar{Y}_\delta\right). \quad (21)$$

Define  $c(t) = X^{(k)\frac{1}{2}}Q(t)\Lambda_{(j)}^{(k)}Q^T(t)X^{(k)\frac{1}{2}}$ ,  $t \in [0, 1]$ . Since  $Q(t) \in \mathbb{O}_n$ , for every  $t \in [0, 1]$ ,  $c(t)$  is a path on  $\mathbb{S}_{++}^n$  from  $Y_{(j)}^{(k)}$  to  $\bar{Y}_\delta$ . Without loss of generality we can assume that  $\delta > d\left(Y_{(j)}^{(k)}, \bar{Y}_\delta\right) \geq d\left(Y_\delta, \bar{Y}_\delta\right)$

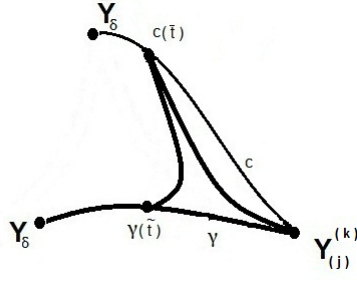


Figure 1: Geodesic segment  $\gamma$ , from  $Y_{(j)}^{(k)}$  to  $Y_\delta$ , and the path  $c$ , from  $Y_{(j)}^{(k)}$  to  $\bar{Y}_\delta$ .

(if this is not true we can choose  $\tilde{t}, \bar{t} \in (0, 1)$  such that  $\delta > d\left(Y_{(j)}^{(k)}, c(\bar{t})\right) \geq d(\gamma(\tilde{t}), c(\bar{t}))$  and we can replace  $Y_\delta$  by  $\gamma(\tilde{t})$  and  $\bar{Y}_\delta$  by  $c(\bar{t})$ . See figure 1). Replacing it in (21), we have that  $d\left(\gamma\left(\frac{1}{2}\right), \bar{Y}_\delta\right) < d\left(Y_{(j)}^{(k)}, \bar{Y}_\delta\right)$ . Hence, for  $\delta > 0$  sufficiently small, the continuity of  $f$  implies that  $(f + \beta^{(k)}f_{X^{(k)}})(\bar{Y}_\delta) < (f + \beta^{(k)}f_{X^{(k)}})\left(Y_{(j)}^{(k)}\right)$  or, equivalently,  $\phi^{(k)}(\lambda_{(j)}^{(k)}, Q_\delta) + \beta^{(k)}\rho_I(\lambda_{(j)}^{(k)}) < \phi^{(k)}(\lambda_{(j)}^{(k)}, Q_{(j)}^{(k)}) + \beta^{(k)}\rho_I(\lambda_{(j)}^{(k)})$ . This implies that  $\phi^{(k)}(\lambda_{(j)}^{(k)}, Q_\delta) < \phi^{(k)}(\lambda_{(j)}^{(k)}, Q_{(j)}^{(k)})$ . It contradicts (ii). To finish, if  $\Lambda_\delta = \Lambda_{(j)}^{(k)}$  then  $\bar{Y}_\delta = Y_\delta$ . So, we conclude that  $Y_\delta$  is the last point of  $c(t)$ . Against, it contradicts (ii). ■

## 4 Inexact version and its complexity

### 4.1 Inexact algorithm

It remarks that  $\text{grad } f(X^{(k+1)}) = -\beta^{(k)}\text{grad } f_{X^{(k)}}(X^{(k+1)})$ , since  $X^{(k+1)}$  is defined as the solution of iteration (4) (remember that  $\text{grad } (f + \beta^{(k)}f_{X^{(k)}})(X^{(k+1)}) = 0$ ). Under Lemmas 2 and 3,  $X^{(k+1)}$  satisfies  $f(X) > f(X^{(k+1)}) + \beta^{(k)}\langle X^{(k+1)\frac{1}{2}} \ln\left(X^{(k+1)^{-\frac{1}{2}} X^{(k)} X^{(k+1)^{-\frac{1}{2}}}\right) X^{(k+1)\frac{1}{2}}, \text{exp}_{X^{(k+1)}}^{-1} X \rangle_{X^{(k+1)}}$ , for every  $X \in \mathbb{S}_{++}^n$ ,  $X \neq X^{(k+1)}$ . Let  $\epsilon^{(k)} \geq 0$ ,  $k = 0, 1, \dots$ , be a sequence of accuracy parameters and  $X^{(0)} \in \mathbb{S}_{++}^n$ . In despite of the exact version, in the inexact algorithm we weaken the late inequality.

**Algorithm 1 (Inexact proximal point algorithm)** Given  $X^{(0)} \in \mathbb{S}_{++}^n$  and  $\beta^{(0)}, \epsilon^{(0)} > 0$ .

1.  $k = 0$ .
2. Computing  $X^{(k+1)} \in \mathbb{S}_{++}^n$  satisfying

$$f(X) \geq f(X^{(k+1)}) + \beta^{(k)}\langle X^{(k+1)\frac{1}{2}} \ln\left(X^{(k+1)^{-\frac{1}{2}} X^{(k)} X^{(k+1)^{-\frac{1}{2}}}\right) X^{(k+1)\frac{1}{2}}, \text{exp}_{X^{(k+1)}}^{-1} X \rangle_{X^{(k+1)}} - \epsilon^{(k)}, \quad (22)$$

for every  $X \in \mathbb{S}_{++}^n$ .

3. Updating  $k$  and returning to 2.

Inequality (22) means that  $\beta^{(k)}X^{(k+1)\frac{1}{2}} \ln\left(X^{(k+1)^{-\frac{1}{2}} X^{(k)} X^{(k+1)^{-\frac{1}{2}}}\right) X^{(k+1)\frac{1}{2}} \in \partial_{\epsilon^{(k)}} f(X^{(k+1)})$ , i. e., it is a  $\epsilon^{(k)}$ -subgradient to  $f$  at  $X^{(k+1)}$ , for  $k = 0, 1, \dots$ . This is the assumption of the inexact version of *SDPProx* algorithm (see  $\epsilon^{(k)}$ -subdifferential relation in [11, p. 475]). For instance, the  $\epsilon$ -subdifferential of a function  $g : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$  at  $Y \in \mathbb{S}_{++}^n$ , represented by  $\partial_\epsilon g(Y)$ , is defined by  $\partial_\epsilon g(Y) = \{S \in \mathbb{S}^n | g(X) \geq g(Y) + \langle S, \text{exp}_Y^{-1} X \rangle_Y - \epsilon, \forall X \in \mathbb{S}_{++}^n\}$ . Note that in the exact version  $\text{grad } f(X^{(k+1)}) = \beta^{(k)}X^{(k+1)\frac{1}{2}} \ln\left(X^{(k+1)^{-\frac{1}{2}} X^{(k)} X^{(k+1)^{-\frac{1}{2}}}\right) X^{(k+1)\frac{1}{2}}$ ,  $k = 0, 1, \dots$ . The exact proximal point algorithm is a particular case of this inexact version, since  $\partial_{\epsilon^{(k)}} f(Y) \supset \partial f(Y) = \{\text{grad } f(Y)\}$ , for every  $Y \in \mathbb{S}_{++}^n$ , where  $\partial f(Y)$  represents the subdifferential of  $f$  at  $Y$ , (if  $\epsilon^{(k)} = 0$ , for every  $k = 0, 1, \dots$ , then  $\partial_{\epsilon^{(k)}} f(X^{(k+1)}) = \partial f(X^{(k+1)}) = \{\text{grad } f(X^{(k+1)})\}$  and inexact version resumes itself to exact proximal point method).

**Theorem 2 (Convergence of the inexact version)** Suppose that  $X^{(k+1)}$  satisfies (22), for every  $k = 0, 1, \dots$ . If the following assertions are held for  $\beta^{(k)}$  and  $\epsilon^{(k)}$

$$\sum_{k=0}^{\infty} \frac{1}{\beta^{(k)}} = +\infty, \quad \sum_{k=0}^{\infty} \epsilon^{(k)} < +\infty, \quad \sum_{k=0}^{\infty} \frac{\epsilon^{(k)}}{\beta^{(k)}} < +\infty, \quad (23)$$

then  $X^{(k)} \rightarrow \mu$ .

**Proof.** It follows directly from Theorem 2 in [11, p. 477]. ■

We still have that  $d^2(X^{(k+1)}, X^{(k)}) = \langle \exp_{X^{(k+1)}}^{-1} X^{(k)}, \exp_{X^{(k+1)}}^{-1} X^{(k)} \rangle_{X^{(k+1)}}$ , where  $\exp_{X^{(k+1)}}^{-1} X^{(k)} = X^{(k+1)\frac{1}{2}} \ln \left( X^{(k+1)-\frac{1}{2}} X^{(k)} X^{(k+1)-\frac{1}{2}} \right) X^{(k+1)\frac{1}{2}}$ . Replacing  $X^{(k)}$  at (22) it follows that  $X^{(k+1)}$  satisfies  $d^2(X^{(k+1)}, X^{(k)}) \leq \frac{\epsilon^{(k)} + f(X^{(k)}) - f(X^{(k+1)})}{\beta^{(k)}}$ ,  $k = 0, 1, \dots$ . On the other hand,  $f(X^{(k)}) \geq \min_{Y \in \mathbb{S}_{++}^n} \{f(Y) + \beta^{(k)} f_{(X^{(k)})}(Y)\} \geq f(X^{(k+1)})$ , since  $f_{X^{(k)}}(X^{(k)}) = 0$  and  $f_{X^{(k)}}(Y) \geq 0$ , for every  $Y \in \mathbb{S}_{++}^n$ . Algorithmically, we repeat iterations (15) and (16) in this order while  $d(Y_{(j+1)}^{(k)}, X^{(k)}) > \sqrt{\frac{\epsilon^{(k)} + f(X^{(k)}) - f(Y_{(j+1)}^{(k)})}{\beta^{(k)}}}$ . When this condition is false we stop both steps and we choose  $Y_{(j+1)}^{(k)}$  to be  $X^{(k+1)}$ .

## 4.2 Complexity

In the inexact version,  $\epsilon^{(k)}$  is previously established for each iteration of the method. The number of iterations necessary to get a  $\epsilon$ -solution for *IKM* problem, for any tolerance  $\epsilon > 0$ , can be estimated under a certain assumption. However, this is not discussed in a practical way here. The following result improves on it.

**Proposition 4** Let  $\{X^{(k)}\}_{k \in \mathbb{N} \cup \{0\}}$  be a sequence generated by (22). Suppose that exists  $\alpha^{(k)} \in (0, \alpha]$ , for any  $\alpha \in (0, 1)$ , such that

$$\frac{\epsilon^{(k)}}{\beta^{(k)}} \leq \alpha^{(k)} d^2(X^{(k+1)}, X^{(k)}), \quad \forall k = 0, 1, \dots, \quad (24)$$

then  $\frac{2}{\beta^{(k)}} \epsilon^{(k)} \left( \frac{1}{\alpha} - 1 \right) \leq d^2(X^{(k)}, \mu)$ , for every  $k = 0, 1, \dots$ .

**Proof.** Since  $\{X^{(k)}\}_{k \in \mathbb{N} \cup \{0\}}$  satisfies (22), Lemma 7 in [11, p. 476] is held. So, we have that

$$d^2(X^{(k+1)}, X) \leq d^2(X^{(k)}, X) - d^2(X^{(k)}, X^{(k+1)}) + \frac{2}{\beta^{(k)}} \left( f(X) - f(X^{(k+1)}) \right) + \frac{2}{\beta^{(k)}} \epsilon^{(k)},$$

for every  $X \in \mathbb{S}_{++}^n$ . In particular, for  $X = \mu$  and taking into account inequality (24), we have that

$$d^2(X^{(k)}, \mu) \geq d^2(X^{(k+1)}, \mu) + d^2(X^{(k)}, X^{(k+1)}) + \frac{2}{\beta^{(k)}} \left( f(X^{(k+1)}) - f(\mu) \right) - \frac{2}{\beta^{(k)}} \epsilon^{(k)}$$

$$\geq d^2(X^{(k+1)}, \mu) + \frac{\epsilon^{(k)}}{\beta^{(k)} \alpha^{(k)}} + \frac{2}{\beta^{(k)}} \left( f(X^{(k+1)}) - f(\mu) \right) - \frac{2}{\beta^{(k)}} \epsilon^{(k)}$$

$$\geq d^2(X^{(k+1)}, \mu) + \frac{2}{\beta^{(k)}} \left( f(X^{(k+1)}) - f(\mu) \right) + \frac{2}{\beta^{(k)}} \epsilon^{(k)} \left( \frac{1}{\alpha} - 1 \right)$$

$$\geq \frac{2}{\beta^{(k)}} \epsilon^{(k)} \left( \frac{1}{\alpha} - 1 \right). \quad \blacksquare$$

**Corollary 4** Let  $\epsilon > 0$  be any tolerance. If  $\beta^{(k)} = \beta$ ,  $\epsilon^{(k)} = \frac{\epsilon^{(0)}}{\tau^k}$ ,  $k = 0, 1, \dots$ , for any  $\beta > 0$  and  $\tau > 1$ , and inequality (24) is held for every  $k$  then at least

$$\lceil \frac{\log(2\epsilon^{(0)}(1-\alpha)) - \log(\beta\alpha\epsilon)}{\log(\tau)} \rceil$$

iterations are necessary to have  $d(X^{(k)}, \mu) \leq \epsilon$ , where  $\lceil t \rceil$  represents the ceiling of  $t$  for any  $t \in \mathbb{R}$ .

**Proof.** Firstly, note that  $\beta^{(k)} = \beta$  and  $\epsilon^{(k)} = \frac{\epsilon^{(0)}}{\tau^k}$ ,  $k = 0, 1, \dots$ , satisfy (23). Put  $d(X^{(k)}, \mu) \leq \epsilon$ . From Proposition 4, we have that  $\tau^k \geq \frac{2\epsilon^{(0)}(1-\alpha)}{\beta\alpha\epsilon}$ . Therefore,

$$k \geq \frac{\log(2\epsilon^{(0)}(1-\alpha)) - \log(\beta\alpha\epsilon)}{\log(\tau)}.$$

## 5 Riemannian gradient flows

### 5.1 Gradient flows on $\mathbb{D}_{++}^n$

Since this work reports to a differentiable Riemannian approach for intrinsic Karcher mean of,  $n \times n$ , SPD matrices, one question emerges naturally: *how can rules for Riemannian gradients of  $\phi^{(k)}$ , as defined in (14), with respect to  $\lambda$  and  $Q$ , be computed?* The answer comes from Riemannian structures of  $\mathbb{S}_{++}^n$ , restricted to  $\mathbb{D}_{++}^n$ , and  $\mathbb{O}_n$ , both discussed in section 2.

**Proposition 5** Set  $Q \in \mathbb{O}_n$ .  $\phi^{(k)}(\cdot, Q) : \mathbb{R}_{++}^n \rightarrow \mathbb{R}$  is geodesic strictly convex and smooth and

$$\text{grad}_\lambda \phi^{(k)}(\lambda, Q) = \lambda \odot \left( \nabla_\lambda \phi^{(k)}(\lambda, Q) \right) \odot \lambda, \quad (25)$$

where  $\nabla_\lambda \phi^{(k)}(\lambda, Q)$  is the Euclidean gradient of  $\phi^{(k)}$ , with respect to  $\lambda$ , at  $(\lambda, Q)$ .

**Proof.** Geodesic strict convexity of  $\phi^{(k)}(\cdot, Q)$  follows from Lemma 2 in [11, p. 472] since  $\mathbb{R}_{++}^n$  is isomorphic to  $\mathbb{D}_{++}^n$ .  $\phi^{(k)}$  is explicitly given by  $\phi^{(k)}(\lambda, Q) = \frac{1}{2} \sum_{i=1}^m \sum_{l=1}^n \ln^2 \theta_l \left( X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q \Lambda Q^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \right)$ ,

where  $\theta_l$  is the  $l^{\text{th}}$  eigenvalue of  $X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q \Lambda Q^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}}$ . Denote by  $\phi^{(k)i}$  the  $i^{\text{th}}$  parcel of  $\phi^{(k)}$ .  $\phi^{(k)i}$  is given by  $\phi^{(k)i}(\lambda, Q) = \sum_{l=1}^n \ln^2 \theta_l \left( X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q \Lambda Q^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \right)$ . Set  $W^{(k)i} = X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q$ .

It is easily seen that  $W^{(k)i}$  is nonsingular,  $W^{(k)iT} = Q^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}}$  and  $W^{(k)iT} \Lambda W^{(k)i}$  is a SPD matrix, since  $\Lambda \in \mathbb{D}_{++}^n$  (see against Theorem 4.5.8 in [12]).  $\phi^{(k)i}$  can be rewritten as  $\phi^{(k)i}(\lambda, Q) = \frac{1}{2} \sum_{l=1}^n \ln^2 \theta_l \left( W^{(k)i} \Lambda W^{(k)iT} \right) = \frac{1}{2} \sum_{l=1}^n \ln^2 \theta_l \left( W^{(k)iT} W^{(k)i} \Lambda \right)$ . Setting  $A^{(k)i} = W^{(k)iT} W^{(k)i}$ , it follows that

$A^{(k)i} \in \mathbb{S}_{++}^n$  and  $\phi^{(k)i}(\lambda, Q) = \frac{1}{2} d^2(A^{(k)i-1}, \Lambda)$ , since  $W^{(k)i}$  is nonsingular and  $A^{(k)i} \Lambda$ , similar to  $A^{(k)i\frac{1}{2}} \Lambda A^{(k)i\frac{1}{2}}$ . So, we conclude that the  $i^{\text{th}}$  parcel of  $\phi^{(k)}$  is geodesic strictly convex and smooth. The Riemannian gradient of  $\phi^{(k)i}$ , with respect to  $\lambda$ , can be computed in alternative way based on Riemannian relation (8). It is easily seen that, in the vector form, the Riemannian gradient of  $\phi^{(k)i}$  is given by

$\text{grad}_\lambda \phi^{(k)i}(\lambda, Q) = \lambda \odot \nabla_\lambda \phi^{(k)i}(\lambda, Q) \odot \lambda$ , where  $\nabla_\lambda \phi^{(k)i}(\lambda, Q)$  is the Euclidean gradient of  $\phi^{(k)i}$ , with respect to  $\lambda$ , at  $(\lambda, Q)$ . Since  $\nabla_\lambda \phi^{(k)}(\lambda, Q) = \sum_{i=1}^n \nabla_\lambda \phi^{(k)i}(\lambda, Q)$ , proposition follows.  $\blacksquare$

$\phi^{(k)i}$  can still be rewritten in a compact form as  $\phi^{(k)i}(\lambda, Q) = \frac{1}{2} \sum_{l=1}^n \ln^2 \theta_l(\lambda, Q)$ , where  $\theta(\lambda, Q) \in \mathbb{R}_{++}^n$

is the vector that contains all eigenvalues of  $X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q \Lambda Q^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}}$  arranged in increasing order.

According to [24, p. 166], a better approximation based in an extrapolation method to the  $l^{th}$  component of  $\nabla_\lambda \phi^{(k)i}(\lambda, Q)$  can be numerically computed by the relation

$$\frac{\phi^{(k)i}(\lambda + he_l, Q) - \phi^{(k)i}(\lambda - he_l, Q)}{h},$$

where  $e_l$  is the  $l^{th}$  vector of the canonic base of  $\mathbb{R}^n$  and  $h > 0$  sufficiently small.

The regularization  $\rho_I(\Lambda) = d^2(I, \Lambda)$  can also be written as a function from  $\mathbb{R}_{++}^n$  to  $\mathbb{R}$ . It is easily seen that  $\rho_I(\lambda) = \frac{1}{2} \sum_{l=1}^n \ln^2 \lambda_l$ . Hence, the Euclidean gradient of  $\rho_I$  is given by  $\nabla \rho_I(\lambda) = \ln(\lambda) \odot \lambda^{-1}$ , where  $\ln(\lambda)$  and  $\lambda^{-1}$  are vectors in  $\mathbb{R}^n$  whose  $l^{th}$  component are  $\ln(\lambda_l)$  and  $\frac{1}{\lambda_l}$ , respectively. Against, applying the Riemannian relation (8), in the vector form, we conclude that

$$\text{grad } \rho_I(\lambda) = \lambda \odot \ln(\lambda). \quad (26)$$

The geodesic rule (6) restricted to  $\mathbb{D}_{++}^n$ , in the vector form, can be rewritten as

$$\gamma(t) = \lambda \odot e^{t\lambda^{-1} \odot s}, \quad (27)$$

for any vector  $s \in \mathbb{R}^n$ , where  $e^x = (e^{x_1}, \dots, e^{x_n})^T$ , for  $x \in \mathbb{R}^n$ .

## 5.2 Gradient flows on $\mathbb{O}_n$

Now, set  $\lambda \in \mathbb{R}_{++}^n$ . We are interested in the Riemannian gradient of  $\phi^{(k)}$ , with respect to  $Q$  on the Riemannian structure of  $\mathbb{O}_n$ . Relation (10) shows how to compute  $\text{grad}_Q \phi^{(k)}(\lambda, Q) \in T_Q \mathbb{O}_n$ . Hence, we need to obtain the rule of the Euclidean gradient of  $\phi^{(k)}$ , with respect to  $Q$ . Since  $\phi^{(k)}(\lambda, Q) = \sum_{i=1}^m \phi^{(k)i}(\lambda, Q)$ , it is sufficient to compute the Euclidean gradient of  $\phi^{(k)i}$ , with respect to  $Q$ . It results

$$\text{that } \nabla_Q \phi^{(k)}(\lambda, Q) = \sum_{i=1}^m \nabla_Q \phi^{(k)i}(\lambda, Q).$$

**Lemma 6** *Let  $A, B : \mathbb{R} \rightarrow \mathbb{R}^{n \times n}$  be differentiable paths. Then,  $[A(t)B(t)]' = A'(t)B(t) + A(t)B'(t)$ .*

**Proof.** Denote  $C(t), D(t)$  e  $E(t) \in \mathbb{R}^{n \times n}$ ,  $t \in \mathbb{R}$  by  $A(t)B(t)$ ,  $A'(t)B(t)$  e  $A(t)B'(t)$ , respectively. Since  $A(t)$  and  $B(t)$  are differentiable paths, all elements  $c_{lq}(t)$ ,  $l, q = 1, \dots, n$  of  $C(t)$  are differentiable scalar functions with respect to  $t$  and

$$c'_{lq}(t) = \left( \sum_{p=1}^n a_{lp}(t) b_{pq}(t) \right)' = \left( \sum_{p=1}^n a'_{lp}(t) b_{pq}(t) \right) + \left( \sum_{p=1}^n a_{lp}(t) b'_{pq}(t) \right) = d_{lq}(t) + e_{lq}(t). \text{ Therefore, } C'(t) = D(t) + E(t) \text{ and the statement follows. } \blacksquare$$

**Proposition 6** *Let  $Q : \mathbb{R} \rightarrow \mathbb{O}^n$  be a differentiable path. Set*

$$R^{(k)i}(t) = X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q(t) \Lambda Q(t)^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}}$$

and  $g(t) = \frac{1}{2} \text{Tr} \left\{ \ln \left( R^{(k)i}(t) \right)^2 \right\}$ . Then the derivative of  $g$ , with respect to  $t$ , is given by

$$g'(t) = 2 \left\langle Q'(t), X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \ln \left( R^{(k)i}(t) \right) \left( R^{(k)i}(t) \right)^{-1} X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q(t) \Lambda \right\rangle_F$$

**Proof.** Since  $X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q(t)$  is nonsingular and  $\Lambda \in \mathbb{D}_{++}^n$ ,  $R^{(k)i}(t)$  is symmetric positive definite, for every  $t \in \mathbb{R}$ . Proposition 2.1 in [17] establishes that  $g'(t) = \text{Tr} \left\{ \ln (R^{(k)i}(t)) (R^{(k)i}(t))^{-1} (R^{(k)i}(t))' \right\}$  and, according to Lema 6,

$$\left( R^{(k)i}(t) \right)' = X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q'(t) \Lambda Q(t)^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} + X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q(t) \Lambda Q'(t)^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}}.$$

This implies that  $g'(t) =$

$$\begin{aligned} &= \text{Tr} \left\{ \ln (R^{(k)i}(t)) (R^{(k)i}(t))^{-1} X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q'(t) \Lambda Q(t)^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \right\} \\ &+ \text{Tr} \left\{ \ln (R^{(k)i}(t)) (R^{(k)i}(t))^{-1} X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q(t) \Lambda Q'(t)^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \right\} \\ &= \text{Tr} \left\{ \Lambda Q(t)^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \ln (R^{(k)i}(t)) (R^{(k)i}(t))^{-1} X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q'(t) \right\} \\ &+ \text{Tr} \left\{ X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \ln (R^{(k)i}(t)) (R^{(k)i}(t))^{-1} X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q(t) \Lambda Q'(t)^T \right\} \\ &= \text{Tr} \left\{ \Lambda Q(t)^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \ln (R^{(k)i}(t)) (R^{(k)i}(t))^{-1} X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q'(t) \right\} \\ &+ \text{Tr} \left\{ Q'(t)^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \ln (R^{(k)i}(t)) (R^{(k)i}(t))^{-1} X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q(t) \Lambda \right\} \\ &= \text{Tr} \left\{ \Lambda Q(t)^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \ln (R^{(k)i}(t)) (R^{(k)i}(t))^{-1} X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q'(t) \right\} \\ &+ \text{Tr} \left\{ \Lambda Q(t)^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} (R^{(k)i}(t))^{-1} \ln (R^{(k)i}(t)) X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q'(t) \right\} \\ &= \left\langle Q'(t), X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} (R^{(k)i}(t))^{-1} \ln (R^{(k)i}(t)) X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q(t) \Lambda \right\rangle_F \\ &+ \left\langle Q'(t), X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \ln (R^{(k)i}(t)) (R^{(k)i}(t))^{-1} X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q(t) \Lambda \right\rangle_F. \end{aligned}$$

The first equality above follows from the substitution of  $(R^{(k)i}(t))'$  in  $g'(t)$ . The second, third and fourth equalities are consequence of similar transformation,  $\text{Tr}(AB) = \text{Tr}(BA)$  and  $\text{Tr}(A) = \text{Tr}(A^T)$ , respectively. The last equality follows from the definition of Frobenius scalar product. Let  $Q^{(k)i} : \mathbb{R} \rightarrow \mathbb{O}_n$  and  $\Lambda^{(k)i} : \mathbb{R} \rightarrow \mathbb{D}_{++}^n$  be paths satisfying  $R^{(k)i}(t) = Q^{(k)i}(t) \Lambda^{(k)i}(t) Q^{(k)i}(t)^T$ . So,  $(R^{(k)i}(t))^{-1} = Q^{(k)i}(t) (\Lambda^{(k)i}(t))^{-1} Q^{(k)i}(t)^T$  and  $\ln (R^{(k)i}(t)) = Q^{(k)i}(t) \ln (\Lambda^{(k)i}(t)) Q^{(k)i}(t)^T$ . Hence  $\ln (R^{(k)i}(t)) (R^{(k)i}(t))^{-1} = Q^{(k)i}(t) \ln (\Lambda^{(k)i}(t)) Q^{(k)i}(t)^T Q^{(k)i}(t) (\Lambda^{(k)i}(t))^{-1} Q^{(k)i}(t)^T$

$$\begin{aligned} &= Q^{(k)i}(t) \ln (\Lambda^{(k)i}(t)) (\Lambda^{(k)i}(t))^{-1} Q^{(k)i}(t)^T \\ &= Q^{(k)i}(t) (\Lambda^{(k)i}(t))^{-1} \ln (\Lambda^{(k)i}(t)) Q^{(k)i}(t)^T \\ &= Q^{(k)i}(t) (\Lambda^{(k)i}(t))^{-1} Q^{(k)i}(t)^T Q^{(k)i}(t) \ln (\Lambda^{(k)i}(t)) Q^{(k)i}(t)^T \\ &= (R^{(k)i}(t))^{-1} \ln (R^{(k)i}(t)). \end{aligned}$$

This means that  $\ln (R^{(k)i}(t))$  and  $(R^{(k)i}(t))^{-1}$  commute and the statement follows.  $\blacksquare$

$\phi^{(k)i}$  can also be written in terms of Frobenius scalar product as

$$\phi^{(k)i}(\lambda, Q) = \frac{1}{2} \text{Tr} \left\{ \ln \left( X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q \Lambda Q^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \right)^2 \right\}.$$

The relation between  $g$  and  $\phi^{(k)i}$  is given by  $g(t) = \phi^{(k)i}(\lambda, Q(t))$ , for any differentiable path  $Q(t)$  on  $\mathbb{O}_n$ . It derives from relation before that  $g'(t) = \langle \nabla_Q \phi^{(k)i}(\lambda, Q(t)), Q'(t) \rangle_F$ . Proposition 6 and the latest equality for  $g'$  imply that  $\nabla_Q \phi^{(k)i}(\lambda, Q) = 2X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \ln (R^{(k)i}(Q)) (R^{(k)i}(Q))^{-1} X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q \Lambda$ , for any  $Q \in \mathbb{O}_n$ , where  $R^{(k)i}(Q) = X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q \Lambda Q^T X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}}$ . Since  $\phi^{(k)}(\lambda, Q) = \sum_{i=1}^m \phi^{(k)i}(\lambda, Q)$ ,

it follows that  $\nabla_Q \phi^{(k)}(\lambda, Q) = 2 \sum_{i=1}^m X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \ln \left( R^{(k)i}(Q) \right) \left( R^{(k)i}(Q) \right)^{-1} X^{i-\frac{1}{2}} X^{(k)\frac{1}{2}} Q \Lambda$ . Replacing  $\left( R^{(k)i}(Q) \right)^{-1}$  in the equality before we have that  $\nabla_Q \phi^{(k)}(\lambda, Q) = 2 \sum_{i=1}^m P^{(k)i}(Q) Q$ , where  $P^{(k)i}(Q) = X^{(k)\frac{1}{2}} X^{i-\frac{1}{2}} \ln \left( R^{(k)i}(Q) \right) X^{i\frac{1}{2}} X^{(k)-\frac{1}{2}}$ . Finally, replacing the late equality in (10) we conclude that the Riemannian gradient of  $\phi^k$  with respect to  $Q$ , at  $T_Q \mathcal{O}_n$ , is given by

$$\text{grad}_Q \phi^{(k)}(\lambda, Q) = \left( \sum_{i=1}^m P^{(k)i} - P^{(k)T} \right) Q, \quad (28)$$

## 6 Simulations and conclusions

### 6.1 Numerical experiments and computational comparisons between proximal point, Richardson-like and geodesic gradient algorithms

In this section are presented numerical experiments with inexact proximal point algorithm. All results are compared with others obtained by geodesic gradient and Richardson-like algorithm. Simulations were made automatically in a workstation Xeon E3 1270V2 3.50 GHz 8MB, Linux and routines were implemented in Matlab, version R2012a.

Sequences  $\{\beta^{(k)}\}_{k \in \mathbb{N} \cup \{0\}}$  and  $\{\epsilon^{(k)}\}_{k \in \mathbb{N} \cup \{0\}}$  were set as in Corollary 4. In addition, we set  $\beta = 10^{-6}$ ,  $\epsilon^{(0)} = 10^{-16}$ ,  $\tau = 1.5625$ ,  $X^{(0)} = Q_{(0)}^{(k)} = \Lambda_{(0)}^{(k)} = I$ ,  $k = 0, 1, \dots$ . We still stopped both steps (15) and (16) when  $d \left( Y_{(j+1)}^{(k)}, X^{(k)} \right) \leq \sqrt{\frac{\epsilon^{(k)}}{\beta^{(k)}}}$ , since it attempts the stopping criteria discussed at the end of Section 4.1.

Tables 1 and 2 present procedures that converts a set  $\mathcal{R} = \{R_1, \dots, R_m\}$  of  $m$ ,  $n \times n$ , nonzero real pseudo-random matrices in a set  $\mathcal{X} = \{X_1, \dots, X_m\}$  of  $m$ ,  $n \times n$ , *SPD* matrices. Pseudo-random matrices  $R_i$ ,  $i = 1, \dots, m$ , were generated in Matlab ambient by syntax  $R_i = \text{rand}(n) - \text{rand}(n)$ . The first table can be found in [3, p. 9] and it generates *SPD* matrices whose eigenvalues belong on the interval  $(0, 1]$ , where  $\lambda_1, \lambda_n$  represent the largest and the smallest eigenvalue of  $X_i$  and  $c_r$ , a condition number previously established.

```

for  $i = 1$  until  $m$  do
   $X_i \leftarrow R_i^T R_i$ ;
   $X_i \leftarrow X_i - \frac{c_r \cdot \lambda_n - \lambda_1}{c_r - 1} \cdot I$ ;
   $X_i \leftarrow \frac{1}{\lambda_n} X_i$ ;
end,

```

Table 1: Normalized matrices set.

```

for  $i = 1$  until  $m$  do
   $X_i \leftarrow \frac{R_i + R_i^T}{2}$ ;
   $X_i \leftarrow e^{X_i}$ ;
end,

```

Table 2: Exponential matrices set.

Since expressions for geodesics and Riemannian gradients in  $\mathcal{O}_n$  and  $\mathbb{D}_{++}^n$  are explicitly known (see rules (9), (25), (26), (27) and (28)), we employ the geodesic gradient method with Armijo's line search as presented in [26] to compute  $\Lambda_{(j+1)}^{(k)}$  and  $Q_{(j+1)}^{(k)}$ . We emphasize that stating an implementable and feasible proximal point method to compute the Riemannian mean of,  $n \times n$ , *SPD* matrices are our objective here. Computational results and comparisons with geodesic gradient and Richardson-like algorithms are only presented to demonstrate the behavior of the method and its applicability.

A bunch of tests was made for each value of  $n$  with sets  $\mathcal{X}_q \subset \mathbb{S}_{++}^n$  ( $|\mathcal{X}_q| = m = 100$ ,  $q = 1, \dots, 20$ ) generated by Table 1, with condition number  $c_r = 10^{2^{r-1}}$ , if  $q \equiv r \pmod{4}$  ( $r = 1, 2, 3$ ), or by Table 2, otherwise. Inexact proximal point method (*PP*), as presented here, Richardson-like (*RL*) and geodesic gradient (*GG*) algorithms, whose implementations can be found in [2], were stopped when the Riemannian distance between two consecutive iterations was smaller than  $\epsilon = 3.16 \times 10^{-9}$ .



Tables 3, 4 and 5 present performances of geodesic gradient, proximal point and Richardson-like algorithms respectively. Notations  $k_{max}$ ,  $k_{min}$ ,  $t_{max}$ ,  $t_{min}$ ,  $EFF_{max}$ ,  $EFF_{min}$  were used to represent the largest and the smallest number of external iterations, CPU time (in seconds) and numerical efficiency respectively. In Table 4, a column  $\lceil \bar{j} \rceil$  is added to represent the ceiling of the mean number  $\bar{j}$  of internal iterations per external iteration in the proximal point algorithm.

Figures 2, 3, 4 and 5 illustrate behaviors of geodesic gradient, proximal point and Richardson-like algorithms on  $\mathbb{S}_{++}^3$ ,  $\mathbb{S}_{++}^5$ ,  $\mathbb{S}_{++}^7$  and  $\mathbb{S}_{++}^{10}$ , with respect to the number of external iterations, CPU time and numerical efficiency.

For  $2 \leq n \leq 5$ ,  $EFF_{PP}$  represents an intermediate case between  $EFF_{GG}$  and  $EFF_{RL}$  and  $PP$  algorithm improves the numerical efficiency of  $GG$  algorithm in the biggest share of the tests, since diagonal positive definite and orthogonal iterates are computed both employing geodesic gradient method. It does not happen for  $6 \leq n \leq 10$ .

We emphasize that many simulations had their automatic execution broken by geodesic gradient or Richardson-like algorithms for sets of matrices generated specially by Table 1 (with condition number  $10^4$ ). These methods employ Cholesky reduction in their code lines and it is unstable for ill-conditioned matrices (instability of Cholesky factorization is discussed with more details in [10, p. 464]). In these cases we restarted other bunch of testes.

$n$	$k_{min}$	$k_{max}$	$t_{min}$ (s)	$t_{max}$ (s)	$EFF_{min}$	$EFF_{max}$
2	1006	1483	71.0905	116.9837	$3.0658e - 05$	$3.1226e - 05$
3	1258	1453	137.6248	162.4914	$3.0977e - 05$	$3.1256e - 05$
4	1271	1482	176.7602	209.8061	$3.0961e - 05$	$3.1278e - 05$
5	1295	1493	224.7501	276.6833	$3.0953e - 05$	$3.1259e - 05$
6	1299	1472	279.1955	335.8208	$3.0961e - 05$	$3.1242e - 05$
7	1303	1504	337.8901	437.7553	$3.0926e - 05$	$3.1262e - 05$
8	1302	1506	424.1944	532.2254	$3.0983e - 05$	$3.1245e - 05$
9	1369	1495	540.8889	648.9021	$3.0916e - 05$	$3.1267e - 05$
10	1312	1496	639.0135	777.5817	$3.0902e - 05$	$3.1260e - 05$

Table 3: Numerical results for geodesic gradient algorithm.

$n$	$k_{min}$	$k_{max}$	$t_{min}$ (s)	$t_{max}$ (s)	$EFF_{min}$	$EFF_{max}$	$\lceil \bar{j} \rceil$
2	2	3	32.9477	186.2839	$2.4214e - 06$	$7.8475e - 06$	8
3	2	10	124.4710	816.0763	$1.4195e - 06$	$1.4626e - 05$	8
4	2	13	238.3308	$2.0330e + 03$	$4.2184e - 06$	$4.4707e - 05$	9
5	2	15	577.7247	$5.1823e + 03$	$5.4695e - 06$	$6.5340e - 05$	9
6	2	14	642.4346	$5.9187e + 03$	$5.6158e - 06$	$7.6468e - 05$	11
7	3	11	$1.1741e + 03$	$1.8143e + 04$	$9.6044e - 06$	$2.2792e - 04$	9
8	3	12	$1.9151e + 03$	$1.2992e + 05$	$2.0295e - 05$	$9.5012e - 04$	10
9	3	10	$2.5564e + 03$	$6.1888e + 04$	$1.0807e - 05$	$4.1031e - 04$	11
10	2	15	$3.5688e + 03$	$8.1416e + 04$	$2.2672e - 05$	$4.1635e - 04$	15

Table 4: Numerical results for proximal point algorithm.

$n$	$k_{min}$	$k_{max}$	$t_{min}$ (s)	$t_{max}$ (s)	$EFF_{min}$	$EFF_{max}$
2	3	18	361.6926	534.3897	$3.2176e - 10$	$4.0360e - 07$
3	5	17	494.8122	621.6392	$4.5054e - 09$	$3.0922e - 07$
4	6	19	533.9189	693.9580	$7.5277e - 09$	$2.8564e - 07$
5	7	19	591.7701	802.3685	$1.2520e - 08$	$2.9640e - 07$
6	8	19	651.6612	921.9018	$2.2759e - 08$	$3.2478e - 07$
7	9	19	732.5926	$1.0920e + 03$	$3.0367e - 08$	$3.2543e - 07$
8	9	20	814.6633	$1.2723e + 03$	$3.0744e - 08$	$3.0353e - 07$
9	9	20	925.5364	$1.4887e + 03$	$3.6233e - 08$	$3.0818e - 07$
10	9	20	$1.0566e + 03$	$1.7371e + 03$	$3.5930e - 08$	$3.4445e - 07$

Table 5: Numerical results for Richardson-like algorithm.

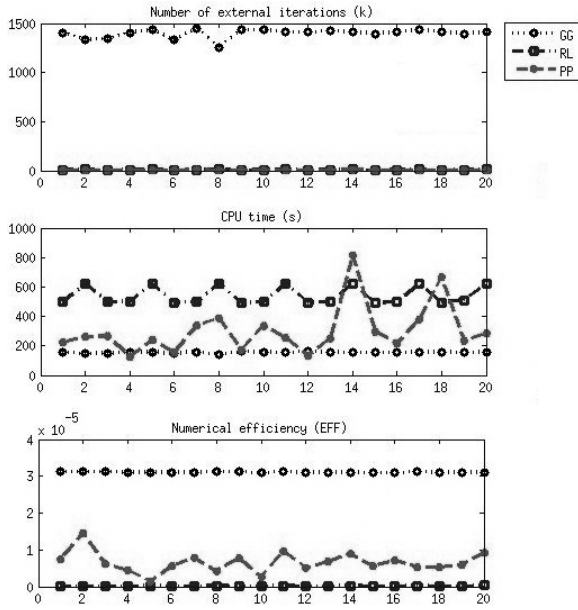


Figure 2:  $n = 3$ .

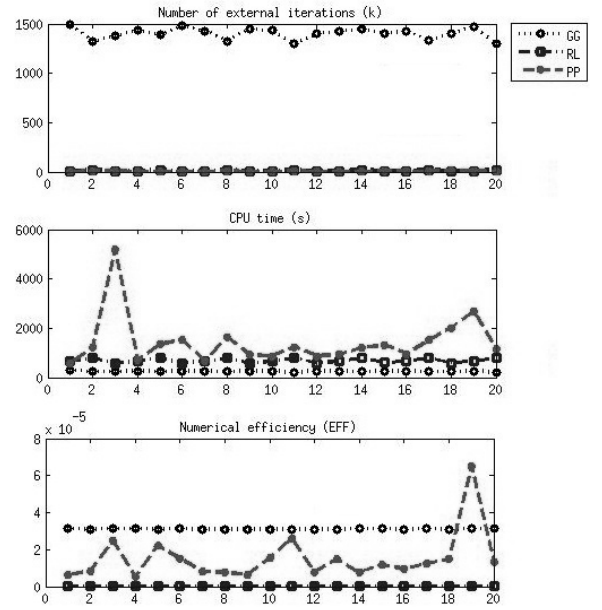


Figure 3:  $n = 5$ .

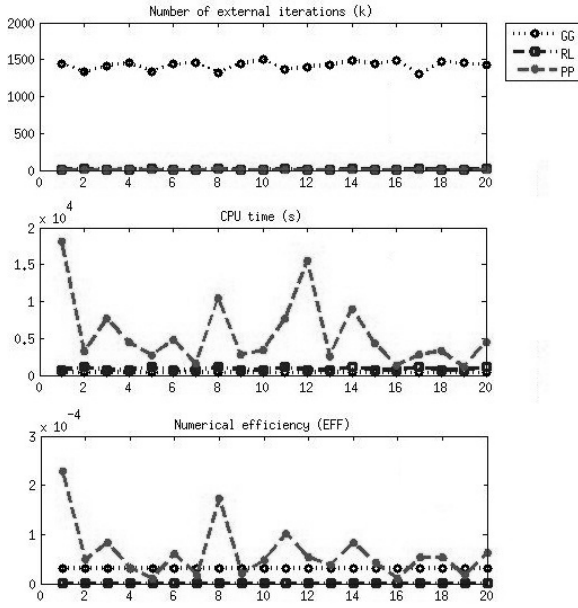


Figure 4:  $n = 7$ .

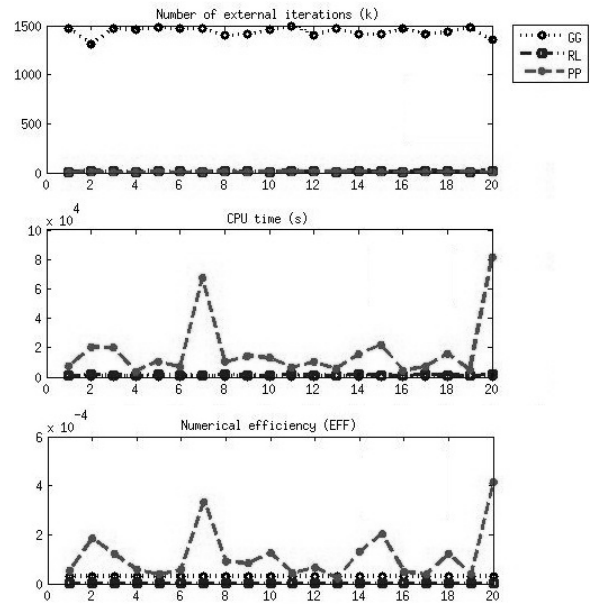


Figure 5:  $n = 10$ .

## 6.2 Conclusions

In this work, we present a feasible proximal point technique to compute the intrinsic Karcher mean of,  $n \times n$ ,  $SPD$  matrices under a strictly Riemannian look. Here, proximal point algorithm works computationally in two phases employing ideas similar to predictor-corrector methods. However, it differs from this class of methods because feasibility of diagonal positive definite and orthogonal steps are held by Riemannian structures of  $\mathbb{S}_{++}^n$ , restricted to  $\mathbb{D}_{++}^n$ , and  $\mathbb{O}_n$  respectively. Solutions for both subproblems generated by our proximal point algorithm are approached employing the geodesic gradient algorithm with Armijo's line search. Unfortunately, (Euclidean or geodesic) gradient methods are not the best choice of algorithm to compute solutions of optimization problems because they spend expensive number of external iterations to return approaches with hard accuracy (or they make tiny

step lengths near the optimum breaking its execution before to get the required accuracy). Other fast and efficient riemannian algorithms to solve both diagonal positive definite and orthogonal subproblems are being investigated. We believe that it will improve the CPU time and numerical efficiency of our method. Since elements of the proximal trajectory are defined as solutions of extended Moreu-Yosida regularizations, whose structures are similar to original *IKM* problem, the methodology employed to determine  $X^{(k+1)}$  can be applied directly to solve it. Still, computational experiments may be made. Real applications, as Riemannian weighted mean of image data in [27], and other geodesic convex problems on *SPD* matrices set will also be investigated under a proximal point view in the future. In addition, proximal point algorithm as presented here can be extended to other domains of positivity where Schur Theorem makes sense. According to [21, p. 189], the hermitian matrices set also represents an example of domain of positivity and spectral theorem is still held for it, where  $\mathbb{O}_n$  is replaced by unitary matrices set  $\mathbb{U}_n$ . For instance, it justifies our conjecture.

**Acknowledgements.** This research was developed in cooperation with researchers from Departamento de Tecnologias e Linguagens of Universidade Federal Rural do Rio de Janeiro and Programa de Engenharia de Sistemas e Computação of Universidade Federal do Rio de Janeiro and it has been supported by **FAPERJ** as part of the research project intituled “**Algoritmo de Ponto proximal em domínios de positividade**”, set in 2012, March. Authors thank **Karine Gomes dos Santos Souza**, a promising computer science student from Departamento de Tecnologias e Linguagens, sponsored by FAPERJ, for partial development of Matlab code to proximal point algorithm and simulations, and FAPERJ for support.

## References

- [1] R. Bhatia, *Positive definite matrices*, Princeton Series in Applied Mathematics, Princeton University Press, New Jersey, 2007.
- [2] D. A. Bini, B. Iannazzo (2010), *The matrix mean toolbox*, <<http://bezout.dm.unipi.it/software/mmttoolbox/>> (retrieved 29.07.11).
- [3] D. A. Bini, B. Iannazzo, *Computing the Karcher mean of symmetric positive definite matrices*, Linear Algebra and its Applications, 438 (2013), pp. 1700-1710.
- [4] R. L. Bishop, B. O’Neill, *Manifolds of negative curvature*, Transactions of the American Mathematical Society, 145 (1969), pp. 1-49.
- [5] C. A. Castaño-moraga, C. Lenglet, R. Deriche, J. Ruiz-Alzola, *A Riemannian approach to anisotropic filtering of tensor fields*, Signal processing, 87 (2007), pp. 263-276.
- [6] O. P. Ferreira, P. R. Oliveira, *Proximal point algorithm on Riemannian manifolds*, Optimization, 51 (2002), pp. 257-270.
- [7] S. Fiori, *Quasi-geodesic neural learning algorithms over the orthogonal group: A tutorial*, Journal of Machine Learning Research, 6 (2005), pp. 743-781.
- [8] S. Fiori, *Learning the Fréchet mean over the manifold of symmetric positive-definite matrices*, Journal of Cognitive Computation, 1 (2009), pp. 279-291.
- [9] P. T. Fletcher, S. Joshi, *Riemannian geometry for the statistical analysis on diffusion tensor data*, Signal processing, 87 (2007), pp. 250-262.
- [10] G. H. Golub, C. Van Loan, *Matrix Computations*, 3<sup>th</sup> edition, The Johns Hopkins University Press, Baltimore, Maryland, 1996.

- [11] R. Gregorio, P. R. Oliveira, *Proximal point algorithm with Schur decomposition on the cone of symmetric semidefinite positive matrices*, Journal of Mathematical Analysis and Applications, 355 (2009), pp. 469-479.
- [12] R. A. Horn, C. R. Johnson, *Matrix Analysis*, 1<sup>st</sup> edition, Cambridge University Press, New York, 1985.
- [13] P. Kingsley, *Introduction to diffusion tensor imaging mathematics: part III. Tensors calculation, noise, simulations and optimization*, Concepts in magnetic resonance part A, 28A (2006), pp. 155-179.
- [14] C. Lenglet, M. Rousson, R. Deriche, O. Faugeras, *Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing*, Journal of Mathematical Imaging and Vision, 25 (2006), pp. 423-444.
- [15] J. H. Manton, *A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups*, in Proceedings of the 8<sup>th</sup> International Conference Control, Automation, Robotics and Vision, 3, 2004, pp. 2211-2216.
- [16] B. Martinet, *Regularisation d'inéquations variationnelles par approximations successives*, Revue Française d'Informatique et Recherche Opérationnelle, 4 (1970), pp. 154-158.
- [17] M. Moakher, *A differential geometry approach to the geometric mean of symmetric positive-definite matrices*, SIAM Journal on Matrix Analysis and Applications, 26 (2005), pp. 735-747.
- [18] X. Pennec, P. Fillard, N. Ayache, *A Riemannian framework for tensor computing*, International Journal of Computer Vision, 66 (2006), pp. 41-66.
- [19] A. P. Papa Quiroz, P. R. Oliveira, *Proximal Point Methods for Quasiconvex and Convex Functions with Bregman Distances on Hadamard Manifolds*, Journal of Convex Analysis, 16 (2009), pp. 49-69.
- [20] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, 14 (1976), pp. 877-898.
- [21] O. S. Rothaus, *Domains of positivity*, Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg, 24 (1960), pp. 189-235.
- [22] T. Sakai, *Riemannian geometry*, Translations of mathematical monographs, 149, American Mathematical Society, Providence, RI, 1996.
- [23] S. T. Smith *Optimization techniques on Riemannian manifolds*, in Hamiltonian and gradient flows, algorithms and control, Fields Institute Communications, 3, American Mathematical Society, Providence, RI, 1994, pp. 113-136.
- [24] J. Stoer, R. Bulirsch, *Introduction to numerical analysis*, 3<sup>th</sup> edition, Springer-Verlag, New York, 2010.
- [25] C. Udriste, *Convex functions and optimization methods on Riemannian manifolds*, Mathematics and Its Applications, Kluwer Academic Publishers, 297, The Netherlands, 1994.
- [26] Y. Yang, *Optimization on Riemannian manifolds*, in Proceedings of the 38<sup>th</sup> conference on decision & control, Phoenix, Arizona, USA, 1999, pp. 888-893.
- [27] F. Zhang, R. E. Hancock, *New Riemannian techniques for directional and tensorial image data*, Pattern recognition, 43 (2010), pp. 1590-1606.