# Averaging symmetric positive-definite matrices*

Xinru Yuan[†] Wen Huang[‡¶] P.-A. Absil[§] K. A. Gallivan[†]

December 7, 2019

**Abstract**

Symmetric positive definite (SPD) matrices have become fundamental computational objects in many areas, such as medical imaging, radar signal processing, and mechanics. For the purpose of denoising, resampling, clustering or classifying data, it is often of interest to average a collection of symmetric positive definite matrices. This paper reviews and proposes different averaging techniques for symmetric positive definite matrices that are based on Riemannian optimization concepts.

## Contents

**1 Introduction** **2**

**2 ALM Properties** **3**

**3 Geodesic Distance Based Averaging Techniques** **4**
3.1 Karcher Mean ($L^2$ Riemannian mean) . . . . . . . . . . . . . 5
3.2 Riemannian Median ($L^1$ Riemannian mean) . . . . . . . . . . 7
3.3 Riemannian Minimax Center ($L^\infty$ Riemannian mean) . . . . 8

*This work was supported by the Fundamental Research Funds for the Central Universities (NO. 20720190060).

[†]Department of Mathematics, Florida State University, Tallahassee FL 32306-4510, USA.

[‡]School of Mathematical Sciences, Xiamen University, P.R.China.

[§]Department of Mathematical Engineering, ICTEAM Institute, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium.

[¶]Corresponding author. E-mail: wen.huang@xmu.edu.cn

**4 Divergence-based Averaging Techniques** **9**
4.1 Divergences . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
4.1.1 The $\alpha$-divergence family . . . . . . . . . . . . . . . . . . 9
4.1.2 Symmetrized divergence . . . . . . . . . . . . . . . . . . 11
4.1.3 The LogDet $\alpha$-divergence . . . . . . . . . . . . . . . . . 11
4.1.4 The LogDet Bregman divergence . . . . . . . . . . . . . 12
4.1.5 The von Neumann $\alpha$-divergence . . . . . . . . . . . . . 13
4.1.6 The von Neumann Bregman divergence . . . . . . . . 13
4.2 Left, Right, and Symmetrized Means Using Divergences . . . 14
4.2.1 The LogDet $\alpha$-divergence . . . . . . . . . . . . . . . . . 15
4.2.2 The LogDet Bregman Divergence . . . . . . . . . . . . . 15
4.2.3 The von Neumann Bregman divergence . . . . . . . . 16
4.3 Divergence-based Median and Minimax Center . . . . . . . . 17

**5 Alternative Metrics on SPD Matrices** **17**

**6 Conclusion** **18**

# 1 Introduction

A symmetric matrix is *positive definite* (SPD) if all its eigenvalues are positive. The set of all $n \times n$ SPD matrices is denoted by

$$\mathcal{S}_{++}^n = \{A \in \mathbb{R}^{n\times n} \mid A = A^T, A \succ 0\},$$

where $A \succ 0$ denotes that all the eigenvalues of $A$ are positive; and an ellipse or an ellipsoid $\{x \in \mathbb{R}^n \mid x^T A x = 1\}$ is used to represent a $2 \times 2$ SPD matrix or larger SPD matrix, see Figure 1.

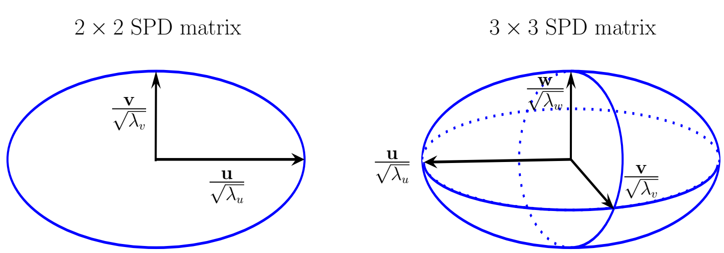


Figure 1: Visualization of an SPD matrix. The axes represent the directions of eigenvectors and the lengths of the axes are the reciprocals of the square roots of the corresponding eigenvalues.

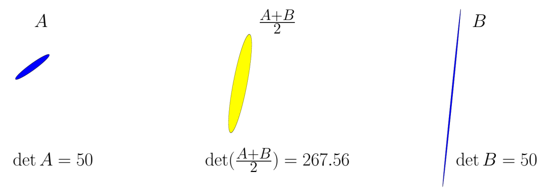


Figure 2: An example of the swelling effect of the arithmetic mean.

SPD matrices have become fundamental computational objects in many areas. For example, they appear as diffusion tensors in medical imaging [25, 32, 60], as data covariance matrices in radar signal processing [15, 42], and as elasticity tensors in elasticity [50]. In these and similar applications, it is often of interest to average or find a central representative for a collection of SPD matrices, e.g., to aggregate several noisy measurements of the same object. Averaging also appears as a subtask in interpolation methods [1] and segmentation [58, 16]. In clustering methods, finding a cluster center as a representative of each cluster is crucial. Hence, it is desirable to find a center that is intrinsically representative and can be computed efficiently.

## 2 ALM Properties

A natural way to average a collection of SPD matrices, $\{A_1, \dots, A_K\}$, is to take their arithmetic mean, i.e., $G(A_1, \dots, A_K) = (A_1 + \cdots + A_K)/K$. However, this is not appropriate in applications where invariance under inversion is required, i.e., $G(A_1, \dots, A_K)^{-1} = G(A_1^{-1}, \dots, A_K^{-1})$. In addition, the arithmetic mean may cause a "swelling effect" that should be avoided in diffusion tensor imaging. Swelling is defined as an increase in the matrix determinant after averaging, see Figure 2 or [32] for more examples. An alternative is to generalize the definition of the geometric mean from scalars to matrices, which yields $G(A_1, \dots, A_K) = (A_1 \dots A_K)^{1/K}$. However, this generalized geometric mean is not invariant under permutation since matrices are not commutative in general. Ando et al. [8] introduced a list of fundamental properties, referred to as the ALM list, that a matrix "geometric" mean should possess:

P1 Consistency with scalars. If $A_1, \dots, A_K$ commute then $G(A_1, \dots, A_K) =$

$(A_1 \cdots A_K)^{1/K}$.

P2 Joint homogeneity. $G(\alpha_1 A_1, \ldots, \alpha_K A_K) = (\alpha_1 \cdots \alpha_K)^{1/K} G(A_1, \ldots, A_K)$.

P3 Permutation invariance. For any permutation $\pi(A_1, \ldots, A_K)$ of $(A_1, \ldots, A_K)$, $G(A_1, \ldots, A_K) = G(\pi(A_1, \ldots, A_K))$.

P4 Monotonicity. If $A_i \geq B_i$ for all $i$, then $G(A_1, \ldots, A_K) \geq G(B_1, \ldots, B_K)$ in the positive semidefinite ordering, i.e., $A \geq B$ iff $A - B \succeq 0$, i.e., $A \geq B$ means that $A - B$ is positive semidefinite (all its eigenvalues are nonnegative).

P5 Continuity from above. If $\{A_1^{(n)}\}, \ldots, \{A_K^{(n)}\}$ are monotonic decreasing sequences (in the positive semidefinite ordering) converging to $A_1, \ldots, A_K$, respectively, then $G(A_1^{(n)}, \ldots, A_K^{(n)})$ converges to $G(A_1, \ldots, A_K)$.

P6 Congruence invariance. $G(S^T A_1 S, \ldots, S^T A_K S) = S^T G(A_1, \ldots, A_K) S$ for any invertible $S$.

P7 Joint concavity. $G(\lambda A_1 + (1-\lambda) B_1, \ldots, \lambda A_K + (1-\lambda) B_K) \geq \lambda G(A_1, \ldots, A_K) + (1 - \lambda) G(B_1, \ldots, B_K)$.

P8 Invariance under inversion. $G(A_1, \ldots, A_K)^{-1} = G(A_1^{-1}, \ldots, A_K^{-1})$.

P9 Determinant identity. $\det G(A_1, \ldots, A_K) = (\det A_1 \cdots \det A_K)^{1/K}$.

These properties are known to be important in numerous applications, e.g. [20, 43, 50]. In the case of $K = 2$, the geometric mean is uniquely defined by the above properties and given by the following expression [17]

$$G(A, B) = A^{\frac{1}{2}} (A^{-\frac{1}{2}} B A^{-\frac{1}{2}})^{\frac{1}{2}} A^{\frac{1}{2}}, \tag{1}$$

where $Z^{\frac{1}{2}}$ for $Z \succ 0$ is the unique SPD matrix such that $Z^{\frac{1}{2}} Z^{\frac{1}{2}} = Z$. However, the ALM properties do not uniquely define a mean for $K \geq 3$. There can be many different definitions of means that satisfy all the properties. The Karcher mean, discussed in Section 3.1, is one of them.

# 3 Geodesic Distance Based Averaging Techniques

Since $\mathcal{S}_{++}^n$ is an open submanifold of the vector space of $n \times n$ symmetric matrices, its tangent space at a point $X$, denoted by $\mathrm{T}_X\, \mathcal{S}_{++}^n$, can be identified with the set of $n \times n$ symmetric matrices. The manifold $\mathcal{S}_{++}^n$ becomes a Riemannian manifold when endowed with the affine-invariant metric,[1] see [58],

[1] The family of Riemanian metrics that satisfy the affine invariance property is described in [34]; see also Section 5. The Riemannian metric (2) is also called the natural metric [31], the trace metric [44], or the Rao–Fisher metric [63].

given by
$$g_X(\xi_X, \eta_X) = \mathrm{trace}(\xi_X X^{-1} \eta_X X^{-1}). \tag{2}$$

The length of a continuously differentiable curve $\gamma : [0, 1] \to \mathcal{M}$ on a Riemannian manifold is

$$\int_0^1 \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))} dt.$$

It is known that, for all $X$ and $Y$ on the Riemannian manifold $\mathcal{S}_{++}^n$ with respect to the metric (2), there is a unique shortest curve such that $\gamma(0) = X$ and $\gamma(1) = Y$. This curve, given by

$$X^{\frac{1}{2}} (X^{-\frac{1}{2}} Y X^{-\frac{1}{2}})^t X^{\frac{1}{2}},$$

is termed a *geodesic*. Its length, given by

$$\delta(X, Y) = \| \log(X^{-1/2} Y X^{-1/2}) \|_{\mathrm{F}},$$

is termed the *geodesic distance* between $X$ and $Y$; see, e.g., [18, Proposition 3] or [58, §3.3].

### 3.1 Karcher Mean ($L^2$ Riemannian mean)

The Karcher mean of $\{A_1, \ldots, A_K\}$, also called the Fréchet mean, the Riemannian barycenter, or the Riemannian center of mass, is defined as the minimizer of the sum of squared distances

$$\mu = \operatorname*{arg\,min}_{X \in \mathcal{S}_{++}^n} F(X), \quad \text{with } F : \mathcal{S}_{++}^{\mathrm{n}} \to \mathbb{R},\ X \mapsto \frac{1}{2K} \sum_{i=1}^{K} \delta^2(X, A_i), \tag{3}$$

where $\delta$ is the geodesic distance associated with metric (2). It is proved in [18, 17] that $F$ is strictly convex and therefore has a unique minimizer. Hence, a point $\mu \in \mathcal{S}_{++}^n$ is a Karcher mean if it is a stationary point of $F$, i.e., $\mathrm{grad}\, F(\mu) = 0$, where $\mathrm{grad}\, F$ denotes the Riemannian gradient of $F$ with respect to the metric (2). The Karcher mean in (3) satisfies all properties in the ALM list [20, 43], and therefore is often used in practice. However, a closed-form solution for problem (3) is not known in general, and for this reason, the Karcher mean is usually computed by iterative methods.

Various methods have been used to compute the Karcher mean of SPD matrices. Most of them resort to the framework of Riemannian optimization (see, e.g., [2]). One exception in [77] resorts to a majorization minimization algorithm. This algorithm is easy to use in the sense that it is a parameter-free algorithm. However, it is usually not as efficient as other Riemannian-optimization-based methods [38]. Several stepsize selection rules have been investigated for the Riemannian steepest descent (RSD) method. A constant stepsize strategy is proposed in [62] and a convergence analysis is given. An adaptive stepsize selection rule based on the explicit expression of the Riemannian Hessian of the cost function $F$ is studied in [61, Algorithm 2], and is shown to be the optimal stepsize for strongly convex cost functions in Euclidean space, see [52, Theorem 2.1.14]. That is, the stepsize is chosen as $\alpha_k = 2/(M_k + L_k)$, where $M_k$ and $L_k$ are the lower and upper bounds on the eigenvalues of the Riemannian Hessian of $F$, respectively. A Riemannian version of the Barzilai-Borwein stepsize (RBB) has been considered in [38]. A version of Newton's method for the Karcher mean computation is also provided in [61]. A Richardson-like iteration is derived and evaluated empirically in [21], and is available in the Matrix Means Toolbox[2]. Yuan has shown in [73] that the Richardson-like iteration is a steepest descent method with stepsize $\alpha_k = 1/L_k$. In [48], a computationally cheap per iteration sequence is analyzed. The method is an incremental gradient algorithm for the cost function (3) based on a shuffled inductive sequence. It is shown that a few iterations gives a matrix that is the best initialization for the state-of-the-art optimization algorithms when compared to commonly-used initial guesses, such as arithmetic-harmonic mean.

A survey of several optimization algorithms for averaging SPD matrices is presented in [39], including Riemannian versions of steepest descent, conjugate gradient, BFGS, and trust-region Newton methods. The authors conclude that the first order methods, steepest descent and conjugate gradient, are the preferred choices for problem (3) in terms of computation time. The benefit of fast convergence of Newton's method and BFGS is nullified by their high computational costs per iteration, especially as the size of the matrices increases. It is also empirically observed in [39] that the Riemannian metric yields much faster convergence for the tested algorithms compared with the induced Euclidean metric, which is given by $g_X(\eta_X, \xi_X) = \mathrm{trace}(\xi_X \eta_X)$.

It is known that a large condition number of the Hessian of the objective function slows down the first order optimization methods. Therefore, a recent paper [74] justifies the observations in [39] by analyzing the condition number of the Hessian in (3). Specifically, it is proven therein that in double precision arithmetic, the condition number of the Hessian of the objective

[2] `http://bezout.dm.unipi.it/software/mmtoolbox/`

function in (3) under the affine-invariance metric (2) is bounded above by a small positive number whereas the condition number of the Hessian under the Euclidean metric is bounded below by a potential large positive number, which linearly depends on the square of the condition number of the minimizer matrix $\mu$. In addition, a limited-memory Riemannian BFGS method is proposed in [75] and empirically shown to be competitive with or superior to other state-of-the-art methods.

## 3.2 Riemannian Median ($L^1$ Riemannian mean)

In the Euclidean space, it is known that the median is preferred to the mean in the presence of outliers due to the robustness of the former and the sensitivity of the latter. This is illustrated in Figure 3, where the mean is dragged towards the outliers lying at the top right corner, while the median appears to be a better estimator of centrality. It is shown in [45] that half of the points must be corrupted in order to corrupt the median.

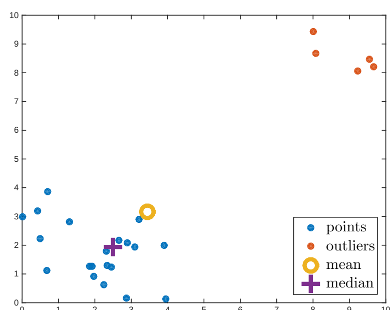


Figure 3: The geometric mean and median in $\mathbb{R}^2$ space.

Given a set of points $\{a_1, \dots, a_K\} \in \mathbb{R}^n$, with the usual Euclidean distance $\|\cdot\|$, the geometric median is defined as the point $m \in \mathbb{R}^n$ minimizing the sum of distance

$$f(x) = \sum_{i=1}^{K} \|x - a_i\|.$$

The geometric median is not available in closed form in general, even for Euclidean points. The geometric median can be computed by an iterative algorithm introduced by Weiszfeld [71], which is essentially an Euclidean steepest descent. Later Ostresh [57] improved Weiszfeld's algorithm and proposed an update iteration with convergence result.

This notion of the geometric median can be extended to the $\mathcal{S}^{\rm n}_{++}$ manifold. Given a set of SPD matrices $\{A_1, \ldots, A_K\}$, their Riemannian median is defined as the minimizer to the sum of distances

$$\mu_1 = \operatorname*{arg\,min}_{X \in \mathcal{S}^{\rm n}_{++}} \sum_{i=1}^{K} \delta(A_i, X), \tag{4}$$

where $\delta(\cdot,\cdot)$ is the geodesic distance. It was proven in [33] that the Riemannian median defined by (4) exists and is unique in the case of a non-positively curved manifold such as $\mathcal{S}^{\rm n}_{++}$ when all the data points $A_i$ do not lie on the same geodesic. Note that the cost function in (4) is not differentiable at the data matrices, i.e., $X = A_i$ for $i = 1, \ldots, K$.

The computation of medians on $\mathcal{S}^{\rm n}_{++}$ has not received as much attention as the mean [33, 23, 73]. Fletcher et al. [33] generalized the Weiszfeld-Ostresh's algorithm to the Riemannian median computation on an arbitrary manifold, and proved that the algorithm converges to the unique solution when it exists. Charfi et al. [23] considered the computation of multiple averaging techniques, including the Riemannian median. An Euclidean steepest descent method and a fixed point algorithm are proposed. However, for the Euclidean steepest descent method, it is not guaranteed that each iterate stays on $\mathcal{S}^{\rm n}_{++}$. no stepsize selection rule is given for the steepest descent method. In [73], Yuan explores Riemannian optimization techniques, in particular smooth and nonsmooth Riemannian quasi-Newton based methods, to compute the Riemannian median, and empirically shows that the limited-memory Riemannian BFGS method is more robust and more efficient than the Riemannian Weiszfeld-Ostresh algorithm.

### 3.3 Riemannian Minimax Center ($L^\infty$ Riemannian mean)

Finding the unique smallest enclosing ball of a finite set of points in a Euclidean space is a fundamental problem in computational geometry and has been explored in e.g., [66, 72, 13, 14, 54]. This can be formulated as finding the minimizer of the cost function $f(x) = \max_{1\leq i \leq K} \|x - a_i\|$. Many data sets from machine learning, medical imaging, or computer vision consist of points on a nonlinear manifold [59, 68]. Therefore, finding the smallest enclosing ball of a collection of points on a manifold is of interest and has been studied in [11]. The center of the smallest enclosing ball is defined to be the $L^\infty$ Riemannian center of mass or the minimax center.

Specifically, given a set of SPD matrices $\{A_1, \ldots, A_K\}$, the minimax center is defined as the point minimizing the maximum geodesic distance $\delta$

to the point set

$$\mu_\infty = \arg\min_{X \in \mathcal{S}^n_{++}} \max_{1 \le i \le K} \delta(A_i, X). \tag{5}$$

In general, there is no known closed form of the solution. In Euclidean space, a fast and simple iterative procedure for solving (5) has been proposed in [13]. The procedure is extended to arbitrary Riemannian manifold in [11] with a study of the convergence rate. The existence and uniqueness of the minimax center defined in (5) have been studied in [3, 4, 11]. The SPD minimax has been used in [9] to denoise tensor images.

The optimization problem in (5) is defined on the Riemannian manifold $\mathcal{S}^n_{++}$. Therefore, Riemannian optimization techniques are natural options for solving this problem. Unlike the cases of the Karcher mean and the median, the solution of (5) usually lies at a non-differentiable point. Therefore, one must utilize nonsmooth optimization techniques on Riemannian manifolds. In [73], Yuan uses the modified Riemannian BFGS method [37] and the subgradient-based Riemannian BFGS method [36] to solve the SPD minimax center problem more efficiently than the state-of-the-art method of Arnaudon and Nielsen [11].

# 4 Divergence-based Averaging Techniques

The averaging techniques based on the geodesic distance provide an attractive approach to averaging a collection of SPD matrices since (i) the approach yields nice geometric interpretations of the optimization problems and (ii) its $L^2$-based Riemannian mean (Karcher mean) satisfies all the desired geometric properties in the ALM list [8].

A divergence is similar to a distance and provides a measure of dissimilarity between two elements. However, in general, it need not satisfy symmetry or the triangle inequality. In recent years, matrix divergences have been of increasing interest due to their simplicity, efficiency and robustness to outliers, e.g., see [70, 10, 69, 23, 27, 55, 28, 7]. The idea of using divergences to define the mean of a collection of SPD matrices has been studied in the literature [50, 51, 26, 65, 64, 24].

## 4.1 Divergences

### 4.1.1 The $\alpha$-divergence family

Let $\varphi : \Omega \to \mathbb{R}$ be a strictly convex and differentiable real-valued function defined on a convex set $\Omega \subset \mathbb{R}^m$. The $\alpha$ divergence family [76] is defined to be

$$\delta^2_{\varphi,\alpha}(x,y) = \frac{4}{1-\alpha^2}[\frac{1-\alpha}{2}\varphi(x) + \frac{1+\alpha}{2}\varphi(y) - \varphi(\frac{1-\alpha}{2}x + \frac{1+\alpha}{2}y)], \quad (6)$$

where $\alpha \in (-1,1)$. The $\alpha$-divergence possesses a dual symmetry with respect to the change $\alpha \to -\alpha$, i.e., $\delta_{\varphi,\alpha}(x,y) = \delta_{\varphi,-\alpha}(y,x)$.

For the values $\alpha = 1$ and $\alpha = -1$, the $\alpha$-divergence is defined by taking the limit as $\alpha \to 1$ and $\alpha \to -1$, i.e.,

$$\delta^2_{\varphi,1}(x,y) = \varphi(x) - \varphi(y) - \langle \nabla\varphi(y), x-y \rangle \text{ and } \delta^2_{\varphi,-1}(x,y) = \delta^2_{\varphi,\mathrm{B}}(y,x). \quad (7)$$

Note that (7) is actually the Bregman divergence defined in [22], denoted by $\delta^2_{\varphi,\mathrm{B}}(x,y)$.

Both the $\alpha$-divergence (6) and the Bregman divergence (7) can be naturally extended to $\mathcal{S}^{\mathrm{n}}_{++}$, e.g., see [50, 24, 53]. Given a strictly convex (in the classical Euclidean sense) and differentiable real-valued function $\phi : \mathcal{S}^{\mathrm{n}}_{++} \to \mathbb{R}$ and $X, Y \in \mathcal{S}^{\mathrm{n}}_{++}$, the $\alpha$-divergence with $-1 < \alpha < 1$ is defined as

$$\delta^2_{\phi,\alpha}(X,Y) = \frac{4}{1-\alpha^2}[\frac{1-\alpha}{2}\phi(X) + \frac{1+\alpha}{2}\phi(Y) - \phi(\frac{1-\alpha}{2}X + \frac{1+\alpha}{2}Y)]. \quad (8)$$

The Bregman divergence, denoted by $\delta^2_{\phi,\mathrm{B}}$, is defined as

$$\delta^2_{\phi,\mathrm{B}}((X,Y) = \phi(X) - \phi(Y) - \langle \nabla\phi(Y), X-Y \rangle, \quad (9)$$

where $\langle X, Y \rangle = \mathrm{tr}(XY)$. Different choices of $\phi$ give different divergences. Commonly used convex functions on $\mathcal{S}^{\mathrm{n}}_{++}$ are [53]:

- quadratic entropy:
$$\phi(X) = \mathrm{tr}(X^T X), \quad (10)$$
- log-determinant (also called Burg) entropy:
$$\phi(X) = -\log\det X, \quad (11)$$
- von Neumann entropy:
$$\phi(X) = \mathrm{tr}(X\log X - X). \quad (12)$$

### 4.1.2 Symmetrized divergence

A divergence is not symmetric in general. There are two common ways to symmetrize a divergence [28]:

- Type 1:
$$\delta^2_{\mathrm{S}\phi}(X,Y) = \frac{1}{2}(\delta^2_\phi(X,Y) + \delta^2_\phi(Y,X)), \tag{13}$$

- Type 2:
$$\delta^2_{\mathrm{S}\phi}(X,Y) = \frac{1}{2}(\delta^2_\phi(X, \frac{X+Y}{2}) + \delta^2_\phi(Y, \frac{X+Y}{2})). \tag{14}$$

### 4.1.3 The LogDet $\alpha$-divergence

When the associated function $\phi(X)$ in (8) is the log-determinant (LogDet) function (11), we get the LogDet $\alpha$-divergence [24]:

$$\delta^2_{\mathrm{LD},\alpha}(X,Y) = \frac{4}{1-\alpha^2} \log \frac{\det(\frac{1-\alpha}{2}X + \frac{1+\alpha}{2}Y)}{[\det(X)]^{\frac{1-\alpha}{2}}[\det(Y)]^{\frac{1+\alpha}{2}}}, \text{ for } -1<\alpha<1. \tag{15}$$

The most frequently mentioned advantage of the LogDet $\alpha$-divergence (15) compared to the geodesic distance $\delta_\mathrm{R}$ is its computational efficiency. The computation of (15) requires three Cholesky factorizations (for $\frac{1-\alpha}{2}X + \frac{1+\alpha}{2}Y$, $X$, and $Y$), while computing the geodesic distance involves eigenvalue decomposition. In addition, the LogDet $\alpha$-divergence enjoys several desired invariance properties [24]:

1. Invariance under congruence transformations
$$\delta^2_{\mathrm{LD},\alpha}(SAS^T, SBS^T) = \delta^2_{\mathrm{LD},\alpha}(A,B) \text{ for any invertible } S. \tag{16}$$

2. Dual-invariance under inversion
$$\delta^2_{\mathrm{LD},\alpha}(A^{-1}, B^{-1}) = \delta^2_{\mathrm{LD},-\alpha}(A,B). \tag{17}$$

3. Dual symmetry
$$\delta^2_{\mathrm{LD},\alpha}(A,B) = \delta^2_{\mathrm{LD},-\alpha}(B,A). \tag{18}$$

The LogDet $\alpha$-divergence (15) is asymmetric except for $\alpha = 0$. But it can be symmetrized using (13) and (14), and the corresponding two symmetric forms of the LogDet $\alpha$-divergence are

$$\delta^2_{\mathrm{S1LD},\alpha}(X,Y) = \frac{2}{1-\alpha^2} \log \frac{\det\left[(\frac{1-\alpha}{2}X + \frac{1+\alpha}{2}Y)(\frac{1-\alpha}{2}Y + \frac{1+\alpha}{2}X)\right]}{\det(XY)}, \quad (19)$$

and

$$\delta^2_{\mathrm{S2LD},\alpha}(X,Y) = \frac{2}{1-\alpha^2} \log \frac{\det\left[(\frac{3-\alpha}{4}X + \frac{1+\alpha}{4}Y)(\frac{3-\alpha}{4}Y + \frac{1+\alpha}{4}X)\right]}{[\det(XY)]^{\frac{1-\alpha}{2}}[\det(\frac{X+Y}{2})]^{1+\alpha}}. \quad (20)$$

The divergence $\delta^2_{\mathrm{LD},0}$ is also called the Stein divergence and is studied in [65, 64]. It is shown in [65] that $\delta^2_{\mathrm{LD},0}$ is the square of a distance function (i.e., $\delta_{\mathrm{LD},0}$ is a distance function in the sense that $\delta_{\mathrm{LD},0}$ is symmetric, nonnegative, definite, and satisfies the triangle inequality), and it shares several common geometric properties with the geodesic distance $\delta^2$, such as P6 (congruence invariance) and P8 (inversion invariance) in the ALM properties, see [65, Table 4.1].

#### 4.1.4 The LogDet Bregman divergence

The LogDet Bregman divergence is defined using $\phi(X) = -\log\det X$, and is given by

$$\delta^2_{\mathrm{LD,B}}(X,Y) = \mathrm{tr}(Y^{-1}X - I) - \log\det(Y^{-1}X). \quad (21)$$

The LogDet Bregman divergence is also called the Kullback-Leibler divergence in [51]. It is easy to verify that the LogDet Bregman divergence is invariant under congruence transformations. In addition, the LogDet Bregman divergence is asymmetric. When it is symmetrized using (13) and (14), we have

$$\delta^2_{\mathrm{S1LD,B}}(X,Y) = \frac{1}{2}\,\mathrm{tr}(Y^{-1}X + X^{-1}Y - 2I), \quad (22)$$

and

$$\delta^2_{\mathrm{S2LD,B}}(X,Y) = \log\det(\frac{X+Y}{2}) - \frac{1}{2}\log\det(XY). \quad (23)$$

Notice that (23) coincides with the LogDet $\alpha$-divergence with $\alpha = 0$. The Type 1 symmetrized LogDet Bregman divergence (22) is also called the Jeffrey divergence (or J-divergence) in [70, 35]. It is easily verified that both (22) and (23) are invariant under congruence and inversion.

#### 4.1.5 The von Neumann $\alpha$-divergence

The von Neumann function $\phi(X) = \operatorname{tr}(X \log X - X)$ arises in quantum mechanics [56]. Its domain is the set of positive semidefinite matrices by using the convention that $0 \log 0 = 0$. The von Neumann $\alpha$-divergence is defined as

$$\begin{aligned}\delta^2_{\mathrm{VN},\alpha}(X,Y) =& \frac{4}{1-\alpha^2} \operatorname{tr}\left\{ \frac{1-\alpha}{2} X \log X + \frac{1+\alpha}{2} Y \log Y \right. \\ & \left. -(\frac{1-\alpha}{2}X + \frac{1+\alpha}{2}Y) \log(\frac{1-\alpha}{2}X + \frac{1+\alpha}{2}Y) \right\}. \end{aligned} \tag{24}$$

From (24), we can verify that the von Neumann $\alpha$-divergence satisfies the following invariance properties:

1. Invariance under rotations

$$\delta^2_{\mathrm{VN},\alpha}(OXO^T, OYO^T) = \delta^2_{\mathrm{VN},\alpha}(X,Y) \text{ for any } O \in \mathrm{SO}(n). \tag{25}$$

2. Dual symmetry

$$\delta^2_{\mathrm{VN},\alpha}(X,Y) = \delta^2_{\mathrm{VN},-\alpha}(Y,X). \tag{26}$$

It is clear from the dual symmetry that the von Neumann divergence is asymmetric except for $\alpha = 0$, which is given by

$$\delta^2_{\mathrm{VN},0}(X,Y) = 4 \operatorname{tr}\{\frac{1}{2} X \log X + \frac{1}{2} Y \log Y - (\frac{X+Y}{2}) \log(\frac{X+Y}{2})\}. \tag{27}$$

We note that the computation of the von Neumann $\alpha$-divergence (24) requires three eigenvalue decompositions, which makes it more expensive than the computation of the geodesic distance $\delta_R$, the LogDet $\alpha$-divergence $\delta^2_{\mathrm{LD},\alpha}$, and the LogDet Bregman divergence $\delta^2_{\mathrm{LD,B}}$. Therefore, we neglect the sided means based on this divergence in Section 4.2.

#### 4.1.6 The von Neumann Bregman divergence

The von Neumann Bregman divergence [53], denoted by $\delta^2_{\mathrm{VN,B}}$, is defined using $\phi(X) = \operatorname{tr}(X \log X - X)$ for the Bregman divergence (9) and is given by

$$\delta^2_{\mathrm{VN,B}}(X,Y) = \operatorname{tr}(X(\log X - \log Y) - X + Y). \tag{28}$$

Note that (28) is referred to as the von Neumann divergence in [40, 29, 53] and the quantum relative entropy in [56]. The von Neumann Bregman divergence (28) is invariant under rotations, and its computation requires two eigenvalue decompositions. It is shown in [29] that (28) is finite if and only if the range of $Y$ contains the range of $X$, i.e., $\mathrm{range}(X) \subseteq \mathrm{range}(Y)$. For this reason, the von Neumann Bregman divergence is often used in low-rank matrix nearness problems, e.g., see [40, 29, 41].

The von Neumann Bregman divergence is not symmetric, and its symmetrized versions are given by

$$\delta^2_{\mathrm{S1VN,B}}(X,Y) = \frac{1}{2}\,\mathrm{tr}(X(\log X - \log Y) + Y(\log Y - \log X)), \tag{29}$$

and

$$\delta^2_{\mathrm{S2VN,B}}(X,Y) = \mathrm{tr}(\frac{1}{2}X\log X + \frac{1}{2}Y\log Y - (\frac{X+Y}{2})\log(\frac{X+Y}{2})). \tag{30}$$

Note that (29) is finite if and only if $\mathrm{range}(X) = \mathrm{range}(Y)$. That is, the Type 1 symmetrized von Neumann Bregman divergence $\delta^2_{\mathrm{S1VN,B}}(X,Y)$ enjoys a range-space preserving property, which is important for the analysis of rank deficient matrices [40]. In addition, we note that the symmetrized von Neumann Bregman divergence (30) coincides with the von Neumann $\alpha$-divergence with $\alpha = 0$, i.e., equation (27).

## 4.2 Left, Right, and Symmetrized Means Using Divergences

Given a divergence function on $\mathcal{S}^{\mathrm{n}}_{++}$, one can define the mean of a collection of SPD matrices $\{A_1, \ldots, A_K\}$ in a way similar to that used for the Karcher mean. Due to the asymmetry of divergence functions, the notion of right mean and left mean are used and coincide if the divergence is symmetric.

**Definition 4.1** *The right mean of a collection of SPD matrices $\{A_1, \ldots, A_K\}$ associated with divergence function $\delta^2_\phi(x,y)$ is defined as the minimizer of the sum of divergences*

$$\mu^{\mathrm{r}} = \operatorname*{arg\,min}_{X \in \mathcal{S}^{\mathrm{n}}_{++}} f(X), \quad \textit{with } f : \mathcal{S}^{\mathrm{n}}_{++} \to \mathbb{R},\ X \mapsto \sum_{i=1}^{K} \delta^2_\phi(A_i, X). \tag{31}$$

**Definition 4.2** *The left mean of a collection of SPD matrices $\{A_1, \ldots, A_K\}$ associated with divergence function $\delta^2_\phi(x,y)$ is defined as the minimizer of the sum of divergences*

$$\mu^{\mathrm{l}} = \operatorname*{arg\,min}_{X \in \mathcal{S}^{\mathrm{n}}_{++}} f(X), \quad \textit{with } f : \mathcal{S}^{\mathrm{n}}_{++} \to \mathbb{R},\ X \mapsto \sum_{i=1}^{K} \delta^2_\phi(X, A_i). \tag{32}$$

**Definition 4.3** *The symmetrized mean of a collection of SPD matrices* $\{A_1, \dots, A_K\}$ *associated with divergence function* $\delta^2_\phi(x,y)$ *is defined as the minimizer of the sum of divergences*

$$\mu^{\mathrm{s}} = \operatorname*{arg\,min}_{X \in \mathcal{S}^{\mathrm{n}}_{++}} f(X), \quad \textit{with } f : \mathcal{S}^{\mathrm{n}}_{++} \to \mathbb{R},\ X \mapsto \sum_{i=1}^{K} \delta^2_{S\phi}(X, A_i). \tag{33}$$

*where* $\delta^2_{S\phi}$ *is defined as* (13) *or* (14).

### 4.2.1 The LogDet $\alpha$-divergence

When $\delta^2_\phi$ is the LogDet $\alpha$-divergence $\delta^2_{\mathrm{LD},\alpha}$, the optimization problems in Definitions 4.1, 4.2 and 4.3 have been studied in [24], where it is proved that the optimization problems have unique minimizers. Sra [65] analyzes the optimization problem for $\alpha = 0$, and proves that $\delta^2_{\mathrm{LD},0}$ is jointly geodesically convex under the affine-invariant metric $g_X(\xi,\eta) = \mathrm{tr}(\xi X^{-1} \eta X^{-1})$ where $\xi, \eta \in \mathrm{T}_X\, \mathcal{S}^{\mathrm{n}}_{++}$. In [73], Yuan extends the result and showed that $\delta^2_{\mathrm{LD},\alpha}$ is jointly geodesically convex for any $-1 < \alpha < 1$. Hence, any local minimum point is also a global minimum point.

A closed-form solution is unknown, except for $K = 2$. Unlike the Karcher mean computation that is extensively tackled by Riemannian optimization methods, the LogDet $\alpha$-divergence based mean is often computed by fixed point algorithms, see [24, 53]. A Euclidean Newton's method is considered in [24] which, however, fails to converge in some numerical experiments. The special case of $\alpha = 0$ is studied in [24] and a fixed point algorithm to compute the divergence-based mean is given and its convergence investigated. This fixed point algorithm is applied to computing the divergence-based mean in [26, 65, 64, 27]. Yuan [73] studies solving the sided mean problem using Riemannian optimization algorithms and explains the fixed point algorithm in [24] in a Riemannian optimization framework. The Riemannian approaches, in particular the limited-memory Riemannian BFGS method, are shown to outperform other state-of-the-art methods for a wide range of problems.

### 4.2.2 The LogDet Bregman Divergence

Means based on the LogDet Bregman divergence have the following closed forms [51, Lemma 17.4.3]:

**Lemma 4.1 ([51, Lemma 17.4.3])** *Let* $\{A_1, \dots, A_K\}$ *be a collection of SPD matrices, let* $\mathcal{A}(A_1, \dots, A_K) = \frac{1}{K} \sum_{i=1}^{K} A_i$ *be their arithmetic mean, let*

$\mathcal{H}(A_1, \ldots, A_K) = K(\sum_{i=1}^{K} A_i^{-1})^{-1}$ *be their harmonic mean, and let* $G(A, B)$ *denote the geometric mean of* $A$ *and* $B$ (1)*.*

1. *The right mean based on* $\delta^2_{\mathrm{LD,B}}$ (21) *is given by the arithmetic mean, i.e.,*

$$\mathcal{A}(A_1, \ldots, A_K) = \arg\min_{X \in \mathcal{S}^{\mathrm{n}}_{++}} \sum_{i=1}^{K} \delta^2_{\mathrm{LD,B}}(A_i, X). \tag{34}$$

2. *The left mean based on* $\delta^2_{\mathrm{LD,B}}$ (21) *is given by the harmonic mean, i.e.,*

$$\mathcal{H}(A_1, \ldots, A_K) = \arg\min_{X \in \mathcal{S}^{\mathrm{n}}_{++}} \sum_{i=1}^{K} \delta^2_{\mathrm{LD,B}}(X, A_i). \tag{35}$$

3. *The symmetric mean based on* $\delta^2_{\mathrm{S1LD,B}}$ (22) *is given by the geometric mean of the arithmetic mean and the harmonic mean, i.e.,*

$$G(\mathcal{A}(A_1, \ldots, A_K), \mathcal{H}(A_1, \ldots, A_K)) = \arg\min_{X \in \mathcal{S}^{\mathrm{n}}_{++}} \sum_{i=1}^{K} \delta^2_{\mathrm{S1LD,B}}(A_i, X). \tag{36}$$

#### 4.2.3 The von Neumann Bregman divergence

Given a collection of SPD matrices $\{A_1, \ldots, A_K\} \in \mathcal{S}^{\mathrm{n}}_{++}$, the right mean $\mu^r$ and left mean $\mu^l$ associated with the von Neumann Bregman divergence are given by, respectively,

$$\mu^{\mathrm{r}} = \arg\min_{X \in \mathcal{S}^{\mathrm{n}}_{++}} \delta^2_{\mathrm{VN,B}}(A_i, X) = \arg\min_{X \in \mathcal{S}^{\mathrm{n}}_{++}} \sum_{i=1}^{K} \mathrm{tr}(A_i \log A_i - A_i \log X - A_i + X) \tag{37}$$

and

$$\mu^{\mathrm{l}} = \arg\min_{X \in \mathcal{S}^{\mathrm{n}}_{++}} \delta^2_{\mathrm{VN,B}}(X, A_i) = \arg\min_{X \in \mathcal{S}^{\mathrm{n}}_{++}} \sum_{i=1}^{K} \mathrm{tr}(X \log X - X \log A_i - X + A_i). \tag{38}$$

In [73], it is pointed out that the left mean based on the von Neumann Bregman divergence has a closed form, which coincides with the Log-Euclidean Fréchet mean in [12]. A closed form of the right mean based on von Neumann Bregman divergence is not known. In addition, no efficient algorithm for computing the right mean currently exists since the closed form of the gradient of $\mathrm{tr}(A_i \log X)$ is not known.

### 4.3 Divergence-based Median and Minimax Center

Similar to the geodesice-distance-based median and minimax center, one can define median and minimax center based on various types of divergences,

$$\text{right median:} \quad \arg\min_{X \in \mathcal{S}_{++}^n} \sum_{i=1}^{K} \delta_{\phi,\alpha}(A_i, X), \tag{39}$$

$$\text{right minimax center:} \quad \arg\min_{X \in \mathcal{S}_{++}^n} \max \delta_{\phi,\alpha}(A_i, X), \tag{40}$$

where $\delta_{\phi,\alpha}$ can be any of the divergences in Section 4.1. The left mean and left minimax center can be defined in a similar way.

In [23], Charfi et al. considered the computation of medians based not only on the geodesic distance, but also on Log-Euclidean distance and the Stein divergence. The Stein divergence median is also studied in [65], and a convergence proof of the fixed point iteration in [23] is given. A median based on the total Kullback-Leibler divergence is proposed in [69], which has a closed form expression. Yuan [73] reviews various types of the divergence-based medians and minimax centers and uses Riemannian optimization techniques to compute those based on the LogDet $\alpha$-divergences. It is shown empirically that Riemannian optimization methods are usually more efficient than other state-of-the-art methods.

## 5 Alternative Metrics on SPD Matrices

Besides the geodesic distance and divergences, there exist other metrics to measure the similarity between two SPD matrices.

**Log-Euclidean metric:** The Log-Euclidean metric proposed in [12] utilizes the observation that the matrix logarithm $\log : \mathcal{S}_{++}^n \to \mathbb{R}^{n\times n}$ is a one-to-one mapping. Therefore, the distance between two SPD matrices $X, Y$ can be defined by

$$\delta_{\text{LogEuc}}(X, Y) = \| \log(X) - \log(Y) \|_F.$$

The Karcher mean defined by this distance has a closed form and coincides with the left mean based on the von Neumann Bregman divergence in Section 4.2.3.

**Wasserstein metric:** The Wasserstein metric defines a general distance between arbitrary probability distributions on a general metric space. Note that the centered multivariate normal distribution $\mathcal{N}(0, X), X \in \mathcal{S}_{++}^n$ is uniquely characterized by $X \in \mathcal{S}_{++}^n$. Therefore, when the Wasserstein metric is used to measure the distance between the multivariate normal distributions with zero mean, it defines a distance metric on $\mathcal{S}_{++}^n$, given by [46]

$$\delta_{\mathrm{Wass}}(X, Y) = \left[\mathrm{tr}(X) + \mathrm{tr}(Y) - 2\,\mathrm{tr}[(X^{\frac{1}{2}} Y X^{\frac{1}{2}})^{\frac{1}{2}}]\right]^{\frac{1}{2}}.$$

The Karcher mean (also called the barycenter) in the Wasserstein space is introduced in [5] and has been used to define the mean on the manifold of $\mathcal{S}_{++}^n$. A fixed point algorithm for computing the Karcher mean of a finite set of probabilities was proposed in [6], and used to find the Karcher mean of SPD matrices. The Wassertein distance can also be interpreted as the geodesic distance in the quotient geometry studied in [19, §4] and [47].

**Affine invariant metric family:** The affine invariance metric family in $\mathcal{S}_{++}^n$ has been studied in [34] and the corresponding geodesic distance is given by

$$\delta_{\mathrm{AIF}}(X, Y) = \left[\frac{\alpha}{4}\,\mathrm{tr}((\log(X^{-1/2} Y X^{-1/2}))^2) + \frac{\beta}{4}(\mathrm{tr}(\log(X^{-1/2} Y X^{-1/2})))^2\right]^{\frac{1}{2}},$$

where $\alpha > 0$ and $\beta > -\alpha/n$. The metric in (2) corresponds to $\alpha = 4$ and $\beta = 0$. In general, the relationship between the Karcher mean based on $\delta_{AIF}$, the choice of parameters of $\alpha$ and $\beta$, and the ALM properties, is not fully understood.

**Other metrics:** Other possibilities include the Bogoliubov-Kubo-Mori [49], the polar affine metric [78] and the broader class of the power Euclidean metrics [30], and the families of balanced metrics introduced in [67].

# 6 Conclusion

In this paper, we have briefly summarized the optimization problems of geodesic-distance-based and divergence-based mean, median and minimax center, and the existing optimization techniques. We have pointed out that the optimization problems in this paper can be nicely solved by Riemannian optimization techniques since the domain $\mathcal{S}_{++}^n$ is a well-studied smooth manifold.

## References

[1] P.-A. Absil, Pierre-Yves Gousenbourger, Paul Striewski, and Benedikt Wirth. Differentiable piecewise-Bézier surfaces on Riemannian manifolds. *SIAM Journal on Imaging Sciences*, 9(4):1788–1828, 2016.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

[3] B. Afsari, R. Tron, and R. Vidal. On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM Journal on Control and Optimization*, 51(3):2230–2260, 2013.

[4] Bijan Afsari. Riemannian $L^p$ center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011.

[5] Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[6] Pedro C. Alvarez-Esteban, E. Del Barrio, J. A. Cuesta-Albertos, and C. Matran. A fixed-point approach to barycenters in Wasserstein space. *Journal of Mathematical Analysis and Applications*, 441(2):744–762, 2016.

[7] Khaled Alyani, Marco Congedo, and Maher Moakher. Diagonality measures of Hermitian positive-definite matrices with application to the approximate joint diagonalization problem. *Linear Algebra and its Applications*, 528(1):290–320, 2017.

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

[9] Jesus Angulo. Structure tensor image filtering using Riemannian $L_1$ and $L_\infty$ center-of-mass. *Image Analysis & Stereology*, 33(2):95–105, 2014.

[10] Ognjen Arandjelovic, Gregory Shakhnarovich, John Fisher, Roberto Cipolla, and Trevor Darrell. Face recognition with image sets using manifold density divergence. In *Proceedings of Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 581–588. IEEE, 2005.

[11] Marc Arnaudon and Frank Nielsen. On approximating the Riemannian 1-center. *Computational Geometry*, 46(1):93–104, 2013.

[12] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic resonance in medicine*, 56(2):411–421, 2006.

[13] Mihai Badoiu and Kenneth L Clarkson. Smaller core-sets for balls. In *Proceedings of Fourteenth ACM-SIAM Symposium on Discrete Algorithms*, 2003.

[14] Mihai Badoiu and Kenneth L. Clarkson. Optimal core-sets for balls. *Computational Geometry*, 40(1):14–22, 2008.

[15] F. Barbaresco. Innovative tools for radar signal processing based on Cartan's geometry of SPD matrices and information geometry. In *Proceedings of IEEE Radar Conference*, pages 1–6, May 2008.

[16] Angelos Barmpoutis, Baba C Vemuri, Timothy M Shepherd, and John R Forder. Tensor splines for interpolation and approximation of DT-MRI with applications to segmentation of isolated rat hippocampi. *IEEE Transactions on Medical Imaging*, 26(11):1537–1546, 2007.

[17] R. Bhatia. *Positive Definite Matrices*. Princeton University Press, 2007.

[18] Rajendra Bhatia and John Holbrook. Riemannian geometry and matrix geometric means. *Linear Algebra and Its Applications*, 413(2-3):594–618, 2006.

[19] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures-Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 2018.

[20] Rajendra Bhatia and Rajeeva L Karandikar. Monotonicity of the matrix geometric mean. *Mathematische Annalen*, 353(4):1453–1467, 2012.

[21] D. A. Bini and B. Iannazzo. Computing the Karcher mean of symmetric positive definite matrices. *Linear Algebra and its Applications*, 438(4):1700–1710, 2013.

[22] Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.

[23] Malek Charfi, Zeineb Chebbi, Maher Moakher, and Baba C Vemuri. Bhattacharyya median of symmetric positive-definite matrices and application to the denoising of diffusion-tensor fields. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pages 1227–1230. IEEE, 2013.

[24] Zeineb Chebbi and Maher Moakher. Means of Hermitian positive-definite matrices based on the log-determinant $\alpha$-divergence function. *Linear Algebra and its Applications*, 436(7):1872–1889, 2012.

[25] Guang Cheng, Hesamoddin Salehian, and Baba Vemuri. Efficient recursive algorithms for computing the mean diffusion tensor and applications to DTI segmentation. *Computer Vision–ECCV 2012*, pages 390–401, 2012.

[26] Anoop Cherian, Suvrit Sra, Arindam Banerjee, and Nikolaos Papanikolopoulos. Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet divergence. In *Proceedings of Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2399–2406. IEEE, 2011.

[27] Anoop Cherian, Suvrit Sra, Arindam Banerjee, and Nikolaos Papanikolopoulos. Jensen-Bregman logdet divergence with application to efficient similarity search for covariance matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2161–2174, 2013.

[28] Andrzej Cichocki, Sergio Cruces, and Shun-ichi Amari. Log-determinant divergences revisited: Alpha-beta and gamma log-det divergences. *Entropy*, 17(5):2988–3034, 2015.

[29] Inderjit S Dhillon and Joel A Tropp. Matrix nearness problems with Bregman divergences. *SIAM Journal on Matrix Analysis and Applications*, 29(4):1120–1146, 2007.

[30] Ian L. Dryden, Xavier Pennec, and Jean Marc Peyrat. Power Euclidean metrics for covariance matrices with application to diffusion tensor imaging. *arXiv:1009.3045v1*, 2010.

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

[32] P. T. Fletcher and S. Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250–262, 2007.

[33] P Thomas Fletcher, Suresh Venkatasubramanian, and Sarang Joshi. The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45(1):S143–S152, 2009.

[34] Wolfgang Forstner and Boudewijn Moonen. *A Metric for Covariance Matrices*. In: Grafarend E. W., Krum F. W., Schwarze V. S. (eds), Geodesy-The Challenge of the 3rd Millennium, Springer Berlin Heidelberg, 2003.

[35] Mehrtash Harandi, Mina Basirat, and Brian C Lovell. Coordinate coding on the Riemannian manifold of symmetric positive-definite matrices for image classification. In *Riemannian Computing in Computer Vision*, pages 345–361. Springer, 2016.

[36] Seyedehsomayeh Hosseini, Wen Huang, and Rohollah Yousefpour. Line search algorithms for locally Lipschitz functions on Riemannian manifolds. *SIAM Journal on Optimization*, 28(1):596–619, 2018.

[37] Wen Huang. *Optimization algorithms on Riemannian manifolds with applications*. PhD thesis, Department of Mathematics, Florida State University, 2014.

[38] Bruno Iannazzo and Margherita Porcelli. The Riemannian Barzilai–Borwein method with nonmonotone line search and the matrix geometric mean computation. *IMA Journal of Numerical Analysis*, 38(1):495–517, 2018.

[39] B. Jeuris, R. Vandebril, and B. Vandereycken. A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electronic Transactions on Numerical Analysis*, 39:379–402, 2012.

[40] Brian Kulis, Mátyás Sustik, and Inderjit Dhillon. Learning low-rank kernel matrices. In *Proceedings of the 23rd international conference on Machine learning*, pages 505–512. ACM, 2006.

[41] Brian Kulis, Mátyás A Sustik, and Inderjit S Dhillon. Low-rank kernel learning with Bregman matrix divergences. *Journal of Machine Learning Research*, 10(Feb):341–376, 2009.

[42] J. Lapuyade-Lahorgue and F. Barbaresco. Radar detection using Siegel distance between autoregressive processes, application to HF and X-band radar. In *Proceedings of IEEE Radar Conference*, pages 1–6, May 2008.

[43] J. Lawson and Y. Lim. Monotonic properties of the least squares mean. *Mathematische Annalen*, 351(2):267–279, 2011.

[44] Jimmie D. Lawson and Yongdo Lim. The Geometric Mean, Matrices, Metrics, and More. *American Mathematical Monthly*, 108(108):797–812, 2001.

[45] Hendrik P Lopuhaa and Peter J Rousseeuw. Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*, pages 229–248, 1991.

[46] Malagó Luigi, Luigi Montrucchio, and Giovanni Pistone. Wasserstein Riemannian geometry of Gaussian densities. *Information Geometry*, 1(2):137–179, 2018.

[47] Estelle Massart and P.-A. Absil. Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices. Technical Report UCL-INMA-2018.06-v2, U.C.Louvain, 2018.

[48] Estelle M Massart, Julien M Hendrickx, and P-A Absil. Matrix geometric means based on shuffled inductive sequences. *Linear Algebra and its Applications*, 542:334–359, 2018.

[49] Peter W. Michor, Dénes Petz, and Attila Andai. On the curvature of a certain Riemannian space of matrices. *Infinite Dimensional Analysis Quantum Probability and Related Topics*, 3(02):199–212, 2000.

[50] M. Moakher. On the averaging of symmetric positive-definite tensors. *Journal of Elasticity*, 82(3):273–296, 2006.

[51] Maher Moakher and Philipp G Batchelor. Symmetric positive-definite matrices: from geometry to applications and visualization. In *Visualization and Processing of Tensor Fields*, pages 285–298. Springer, 2006.

[52] Yurii Nesterov. Introductory lectures on convex programming volume I: Basic course. *Lecture notes*, 1998.

[53] Frank Nielsen, Meizhu Liu, Xiaojing Ye, and Baba C Vemuri. Jensen divergence based SPD matrix means and applications. In *Proceedings of 21st International Conference on Pattern Recognition (ICPR)*, pages 2841–2844. IEEE, 2012.

[54] Frank Nielsen and Richard Nock. Approximating smallest enclosing balls with applications to machine learning. *International Journal of Computational Geometry and Applications*, 19(05):389–414, 2009.

[55] Frank Nielsen and Richard Nock. Total Jensen divergences: definition, properties and clustering. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 2016–2020. IEEE, 2015.

[56] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

[57] Lawrence M Ostresh Jr. On the convergence of a class of iterative methods for solving the Weber location problem. *Operations Research*, 26(4):597–609, 1978.

[58] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

[59] Xavier Pennec. *Statistical Computing on Manifolds: From Riemannian Geometry to Computational Anatomy*, pages 347–386. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[60] Y. Rathi, A. Tannenbaum, and O. Michailovich. Segmenting images on the tensor manifold. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[61] Q. Rentmeesters. *Algorithms for data fitting on some common homogeneous spaces*. PhD thesis, Universite catholiqué de Louvain, 2013.

[62] Q. Rentmeesters and P.-A. Absil. Algorithm comparison for Karcher mean computation of rotation matrices and diffusion tensors. In *19th European Signal Processing Conference*, pages 2229–2233, Aug 2011.

[63] Salem Said, Lionel Bombrun, Yannick Berthoumieu, and Jonathan H. Manton. Riemannian Gaussian Distributions on the Space of Symmetric Positive Definite Matrices. *IEEE Transactions on Information Theory*, 63(4):2153–2170, 2017.

[64] Suvrit Sra. A new metric on the manifold of kernel matrices with application to matrix geometric means. In *Advances in Neural Information Processing Systems*, pages 144–152, 2012.

[65] Suvrit Sra. Positive definite matrices and the S-divergence. *Proceedings of the American Mathematical Society*, 144(7):2787–2797, 2015.

[66] J. J. Sylvester. A question in the geometry of situation,. *Q. J. Math.*, 1:17, 1857.

[67] Yann Thanwerdas and Xavier Pennec. Exploration of balanced metrics on symmetric positive definite matrices. In Frank Nielsen and Frédéric Barbaresco, editors, *Geometric Science of Information*, pages 484–493, Cham, 2019. Springer International Publishing.

[68] Pavan Turaga, Ashok Veeraraghavan, Anuj Srivastava, and Rama Chellappa. Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11):2273–86, 2011.

[69] Baba C Vemuri, Meizhu Liu, Shun-Ichi Amari, and Frank Nielsen. Total Bregman divergence and its applications to DTI analysis. *IEEE Transactions on medical imaging*, 30(2):475–483, 2011.

[70] Zhizhou Wang and Baba C Vemuri. An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation. In *Proceedings of Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2004.

[71] Endre Weiszfeld. Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohoku Mathematical Journal, First Series*, 43:355–386, 1937.

[72] E. Welzl. Smallest enclosing disks (balls and ellipsoids). *New Results and New Trends in Computer Science*, 1991.

[73] Xinru Yuan. *Riemannian optimization methods for averaging symmetric positive definite matrices*. PhD thesis, Department of Mathematics, Florida State University, 2018.

[74] Xinru Yuan, Wen Huang, P.-A. Absil, and K. A. Gallivan. Computing the matrix geometric mean: Riemannian vs Euclidean conditioning, implementation techniques, and a Riemannian BFGS method. Technical Report UCL-INMA-2019.05, U.C.Louvain, 2019. https://www.math.fsu.edu/ whuang2/papers/CMGM.htm.

[75] Xinru Yuan, Wen Huang, P.-A. Absil, and Kyle A. Gallivan. A Riemannian limited-memory BFGS algorithm for computing the matrix geometric mean. *Procedia Computer Science*, 80:2147–2157, 2016.

[76] Jun Zhang. Divergence function, duality, and convex analysis. *Neural Computation*, 16(1):159–195, 2004.

[77] T. Zhang. A MAJORIZATION-MINIMIZATION ALGORITHM FOR COMPUTING THE KARCHER MEAN OF POSITIVE DEFINITE MATRICES. *SIAM Journal on Matrix Analysis and Applications*, 38(2):387–400, 2017.

[78] Zhengwu Zhang, Jingyong Su, Eric Klassen, Huiling Le, and Anuj Srivastava. Rate-invariant analysis of covariance trajectories. *Journal of Mathematical Imaging and Vision*, 60(8):1306–1323, 2018.