

RIEMANNIAN GEOMETRY OF SYMMETRIC POSITIVE DEFINITE MATRICES VIA CHOLESKY DECOMPOSITION*

ZHENHUA LIN[†]

Abstract. We present a new Riemannian metric, termed Log-Cholesky metric, on the manifold of symmetric positive definite (SPD) matrices via Cholesky decomposition. We first construct a Lie group structure and a bi-invariant metric on Cholesky space, the collection of lower triangular matrices whose diagonal elements are all positive. Such group structure and metric are then pushed forward to the space of SPD matrices via the inverse of Cholesky decomposition that is a bijective map between Cholesky space and SPD matrix space. This new Riemannian metric and Lie group structure fully circumvent swelling effect, in the sense that the determinant of the Fréchet average of a set of SPD matrices under the presented metric, called Log-Cholesky average, is between the minimum and the maximum of the determinants of the original SPD matrices. Comparing to existing metrics such as the affine-invariant metric and Log-Euclidean metric, the presented metric is simpler, more computationally efficient, and numerically stabler. In particular, parallel transport along geodesics under Log-Cholesky metric is given in a closed and easy-to-compute form.

Key words. Fréchet mean, symmetric positive definite matrix, Lie group, bi-invariant metric, parallel transport, Cholesky decomposition, lower triangular matrix

AMS subject classifications. 47A64, 26E60, 53C35, 22E99, 32F45, 53C22, 15A22

DOI. 10.1137/18M1221084

1. Introduction. Symmetric positive definite (SPD) matrices emerge in vast scientific applications such as computer vision [9, 35], elasticity [18, 31], signal processing [3, 21], medical imaging [11, 13, 14, 27, 39], and neuroscience [15]. A concrete example is analysis of functional connectivity between brain regions. Such connectivity is often characterized by the covariance of blood-oxygen-level dependent signals [22] generated by brain activities from different regions. The covariance is mathematically defined by a covariance matrix which is an SPD matrix. Another application is diffusion tensor imaging [25], which is extensively used to obtain high-resolution information of internal structures of certain tissues or organs, such as hearts and brains. For each tissue voxel, there is a 3×3 SPD matrix to describe the shape of local diffusion. Such information has clinical applications; for example, it can be used to discover pathological area surrounded by healthy tissues.

The space of SPD matrices of a fixed dimension m , denoted by \mathcal{S}_m^+ in this article, is a convex smooth submanifold of the Euclidean space $\mathbb{R}^{m(m+1)/2}$. The inherited Euclidean metric further turns \mathcal{S}_m^+ into a Riemannian manifold. However, as pointed out in [4], this classic metric is not adequate in many applications for two reasons. First, the distance between SPD matrices and symmetric matrices with zero or negative eigenvalues is finite, which implies that, in the context of diffusion tensor imaging, small diffusion is more likely than large diffusion. Second, the Euclidean average of SPD matrices suffers from swelling effect, i.e., the determinant of the average is larger than any of the original determinants. When SPD matrices are covariance matrices, as in the application of diffusion tensor imaging, determinants correspond to overall dispersion of diffusion. Inflated determinants amount to extra diffusion that is artificially introduced in computation.

*Received by the editors October 16, 2018; accepted for publication (in revised form) by P.-A. Absil April 11, 2019; published electronically November 19, 2019.

<https://doi.org/10.1137/18M1221084>

[†]National University of Singapore, Singapore, 117546 (stalz@nus.edu.sg).

To circumvent the problems of the Euclidean metric for SPD matrices, various metrics have been introduced in the literature, such as the affine-invariant metric [30, 34] and the Log-Euclidean metric [4]. These metrics keep symmetric matrices with some nonpositive eigenvalues at an infinite distance away from SPD matrices and are not subject to swelling effect. In addition, the Log-Euclidean framework features a closed form of the Fréchet average of SPD matrices. It also turns \mathcal{S}_m^+ into a Lie group endowed with a bi-invariant metric. However, computation of Riemannian exponential and logarithmic maps requires evaluating a series of an infinite number of terms; see (2.1) and (3.4) in [4]. Comparing to the Log-Euclidean metric, the affine-invariant one not only possesses easy-to-compute exponential and logarithmic maps but also enjoys a closed form for parallel transport along geodesics; see Lemma 3 of [36]. However, to the best of our knowledge, no closed form is found for the Fréchet average of SPD matrices under the affine-invariant metric. The Fréchet average of SPD matrices is also studied in the literature for distance functions or Riemannian metrics arising from perspectives other than swelling effect, such as the Bures–Wasserstein metric that is related to the theory of optimal transport [7] and the S-divergence studied in both [10] and [37]. Other related works include Riemannian geometry for positive semidefinite matrices [38, 28] and Riemannian structure for correlation matrices [17].

In addition to the above Riemannian frameworks, it is also common to approach SPD matrices via Cholesky decomposition in practice for efficient computation, such as [12, 32, 39]. Distance on SPD matrices based on Cholesky decomposition has also been explored in the literature. For example, in [11] the distance between two SPD matrices P_1 and P_2 with Cholesky decomposition $P_1 = L_1 L_1^\top$ and $P_2 = L_2 L_2^\top$ is defined by $\|L_1 - L_2\|_F$, where each of L_1 and L_2 is a lower triangular matrix whose diagonal elements are positive, and $\|\cdot\|_F$ denotes Frobenius matrix norm. Although this distance is simple and easy to compute, it suffers from swelling effect, as demonstrated by the following example.

Example 1. One first notes that, under the Cholesky distance, the geodesic interpolation between P_1 and P_2 is given by $P_\rho := \{\rho L_1 + (1 - \rho)L_2\}\{\rho L_1 + (1 - \rho)L_2\}^\top$ for $\rho \in [0, 1]$. For any $\epsilon > 0$, consider matrices

$$P_1 = \begin{pmatrix} \epsilon^2 & 0 \\ 0 & 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1 & 0 \\ 0 & \epsilon^2 \end{pmatrix}, \quad L_1 = \begin{pmatrix} \epsilon & 0 \\ 0 & 1 \end{pmatrix}, \quad L_2 = \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix}.$$

It is clear that $L_1 L_1^\top = P_1$ and $L_2 L_2^\top = P_2$. When $\rho = 1/2$,

$$P_\rho = \frac{1}{4}(L_1 + L_2)(L_1 + L_2)^\top = \begin{pmatrix} \frac{(1+\epsilon)^2}{4} & 0 \\ 0 & \frac{(1+\epsilon)^2}{4} \end{pmatrix},$$

whose determinant is $\det(P_\rho) = (1 + \epsilon)^4/16$. However, $\det(P_1) = \det(P_2) = \epsilon^2 < (1 + \epsilon)^4/16$, or equivalently, $\max\{\det(P_1), \det(P_2)\} < \det(P_\rho)$, whenever $\epsilon \neq 1$.

In this work, we propose a new Riemannian metric on SPD matrices via Cholesky decomposition. The basic idea is to introduce a new metric for the space of lower triangular matrices with positive diagonal elements and then push it forward to the space of SPD matrices via Cholesky decomposition. The metric, termed *Log-Cholesky metric*, has the advantages of the aforementioned affine-invariant metric, Log-Euclidean metric and Cholesky distance. First, it is as simple as the Cholesky distance but not subject to swelling effect. Second, like the Log-Euclidean metric, the presented metric enjoys Lie group bi-invariance as well as a closed form for the Log-Cholesky

average of SPD matrices. This bi-invariant Lie group structure seems not shared by the aforementioned works other than [4] in the literature. Third, it features simple and easy-to-compute expressions for Riemannian exponential and logarithmic maps, in contrast with the Log-Euclidean metric. Finally, like the affine-invariant metric, the expression for parallel transport along geodesics is simple and easy-to-compute under the presented metric. Parallel transport is important in applications like regression methods on Riemannian manifolds, such as [36, 40, 41].

It is noted that Cholesky decomposition is also explored in [17] for a Riemannian geometry of correlation matrices with rank no larger than a fixed bound. Despite certain similarity in the use of Cholesky decomposition, this work is fundamentally different from ours. First, it studies correlation matrices rather than SPD matrices. For a correlation matrix, its diagonal elements are restricted to be one. Second, the Riemannian structures considered in [17] and our work are different. For example, the so-called Cholesky manifold in [17] is a Riemannian submanifold of a Euclidean space, while our Riemannian manifold to be proposed is not. Finally, Cholesky decomposition is utilized in [17] as a way to parameterize correlation matrices, rather than push forward a new manifold structure to correlation matrices.

We structure the rest of this article as follows. Some notations and basic properties of lower triangular and SPD matrices are collected in section 2. In section 3, we introduce a new Lie group structure on SPD matrices and define the Log-Cholesky metric on the group. Basic features such as Riemannian exponential/logarithmic maps, geodesics, and parallel transport are also characterized. Section 4 is devoted to the Log-Cholesky mean/average of distributions on SPD matrices. We then conclude the article in section 5.

2. Lower triangular matrices and SPD matrices. We start with introducing some notations and recalling some basic properties of lower triangular and SPD matrices. Cholesky decomposition is then shown to be a diffeomorphism between lower triangular matrix manifolds and SPD manifolds. This result serves as a cornerstone of our development: it enables us to push forward a Riemannian metric defined on the space of triangular matrices to the space of SPD matrices.

2.1. Notations and basic properties. Throughout this paper, m is a fixed positive integer that represents the dimension of matrices under consideration. For a matrix A , we use A_{ij} or $A(i, j)$ to denote its element on the i th row and j th column. The notation $[A]$ denotes an $m \times m$ matrix whose (i, j) element is A_{ij} if $i > j$ and is zero otherwise, while $\mathbb{D}(A)$ denotes an $m \times m$ diagonal matrix whose (i, i) element is A_{ii} . In other words, $[A]$ is the strictly lower triangular part, while $\mathbb{D}(A)$ is the diagonal part of A . The trace of a matrix A is denoted by $\text{tr}(A)$, and the determinant is denoted by $\det(A)$. For two square matrices A and B , $\langle A, B \rangle_F := \sum_{ij} A_{ij} B_{ij}$ denotes the Frobenius inner product between them, and the induced norm is denoted by $\|A\|_F := \langle A, A \rangle_F^{1/2}$.

The matrix exponential map of a real matrix is defined by $\exp(A) = \sum_{k=0}^{\infty} A^k / k!$, and its inverse, the matrix logarithm, whenever it exists and is real, is denoted by $\log(A)$. It is noted that the exponential of a lower triangular matrix is also lower triangular. In addition, the matrix exponential of a diagonal matrix can be obtained by applying the exponential function to each diagonal element. The matrix logarithm of a diagonal matrix with positive diagonal elements can be computed in a similar way. Thus, the matrix exponential/logarithmic map of a diagonal matrix is diagonal.

The space of $m \times m$ lower triangular matrices is denoted by \mathcal{L} , and the subset of \mathcal{L} whose diagonal elements are all positive is denoted by \mathcal{L}_+ . It is straightforward to check the following properties of lower triangular matrices.

- $X = \lfloor X \rfloor + \mathbb{D}(X)$ for $X \in \mathcal{L}$.
- $X_1 + X_2 \in \mathcal{L}$ and $X_1 X_2 \in \mathcal{L}$ for $X_1, X_2 \in \mathcal{L}$.
- $L_1 + L_2 \in \mathcal{L}_+$ and $L_1 L_2 \in \mathcal{L}_+$ if $L_1, L_2 \in \mathcal{L}_+$.
- If $L \in \mathcal{L}_+$, then the inverse L^{-1} exists and belongs to \mathcal{L}_+ .
- For $X_1, X_2 \in \mathcal{L}$, $\mathbb{D}(X_1 + X_2) = \mathbb{D}(X_1) + \mathbb{D}(X_2)$ and $\mathbb{D}(X_1 X_2) = \mathbb{D}(X_1) \mathbb{D}(X_2)$.
- $\mathbb{D}(L^{-1}) = \mathbb{D}(L)^{-1}$ for $L \in \mathcal{L}_+$.
- $\det(X) = \prod_{j=1}^m X_{jj}$ for $X \in \mathcal{L}$.

These properties show that both \mathcal{L} and \mathcal{L}_+ are closed under matrix addition and multiplication and that the operator \mathbb{D} interacts well with these operations.

Recall that \mathcal{S}_m^+ is defined as the collection of $m \times m$ SPD matrices. We denote the space of $m \times m$ symmetric space by \mathcal{S}_m . Symmetric matrices and SPD matrices possess numerous algebraic and analytic properties that are well documented in [6]. Below are some of them to be used in what follows.

- All eigenvalues $\lambda_1, \dots, \lambda_m$ of an SPD P are positive, and $\det(P) = \prod_{j=1}^m \lambda_j$. Therefore, the determinant of an SPD matrix is positive.
- For any invertible matrix X , the matrix XX^\top is an SPD matrix.
- $\exp(S)$ is an SPD matrix for a symmetric matrix S , while $\log(P)$ is a symmetric metric for an SPD matrix P .
- Diagonal elements of an SPD matrix are all positive. This can be seen from the fact that $P_{jj} = e_j^\top P e_j > 0$ for $P \in \mathcal{S}_m^+$, where e_j is the unit vector with 1 at the j th coordinate and 0 elsewhere.

2.2. Cholesky decomposition. Cholesky decomposition, named after André-Louis Cholesky, represents a real $m \times m$ SPD matrix P as a product of a lower triangular matrix L and its transpose, i.e., $P = LL^\top$. If the diagonal elements of L are restricted to be positive, then the decomposition is unique according to Theorem 4.2.5 of [16]. Such lower triangular matrix, denoted by $\mathcal{L}(P)$, is called the *Cholesky factor* of P . Since in addition $L = \mathcal{L}(LL^\top)$ for each $L \in \mathcal{L}_+$, the map $\mathcal{L} : \mathcal{S}_m^+ \rightarrow \mathcal{L}_+$ is bijective. In other words, there is one-to-one correspondence between SPD matrices and lower triangular matrices whose diagonal elements are all positive.

The space \mathcal{L}_+ , called the *Cholesky space* in this paper, is a smooth submanifold of \mathcal{L} that is identified with the Euclidean space $\mathbb{R}^{m(m+1)/2}$. Similarly, the space \mathcal{S}_m^+ of SPD matrices is a smooth submanifold of the space \mathcal{S}_m of symmetric matrices identified with vectors in $\mathbb{R}^{m(m+1)/2}$. As a manifold map between smooth manifolds \mathcal{L}_+ and \mathcal{S}_m^+ , the map \mathcal{L} is indeed a diffeomorphism. This fact will be explored to endow \mathcal{S}_m^+ with a new Riemannian metric that to be presented in subsection 3.2.

PROPOSITION 2. *The Cholesky map \mathcal{L} is a diffeomorphism between smooth manifolds \mathcal{L}_+ and \mathcal{S}_m^+ .*

Proof. We have argued that \mathcal{L} is a bijection. To see that it is also smooth, for $P = LL^\top$ with $L \in \mathcal{L}_+$, we write

$$\begin{aligned} \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1m} \\ P_{21} & P_{22} & \cdots & P_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ P_{m1} & P_{m2} & \cdots & P_{mm} \end{pmatrix} &= \begin{pmatrix} L_{11} & 0 & \cdots & 0 \\ L_{21} & L_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ L_{m1} & L_{m2} & \cdots & L_{mm} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & \cdots & L_{m1} \\ 0 & L_{22} & \cdots & L_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & L_{mm} \end{pmatrix} \\ &= \begin{pmatrix} L_{11}^2 & L_{21}L_{11} & \cdots & L_{m1}L_{11} \\ L_{21}L_{11} & L_{21}^2 + L_{22}^2 & \cdots & L_{m1}L_{21} + L_{m2}L_{22} \\ \vdots & \vdots & \ddots & \vdots \\ L_{m1}L_{11} & L_{m1}L_{21} + L_{m2}L_{22} & \cdots & \sum_{k=1}^m L_{mk}^2 \end{pmatrix}, \end{aligned}$$

from which we deduce that

$$(2.1) \quad \begin{cases} L_{ii} = \sqrt{P_{ii} - \sum_{k=1}^{i-1} L_{ik}^2}, \\ L_{ij} = \frac{1}{L_{jj}} \left(P_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{jk} \right) \quad \text{for } i > j. \end{cases}$$

The existence of a unique Cholesky factor for every SPD matrix suggests that $P_{ii} - \sum_{k=1}^{i-1} L_{ik}^2 > 0$ for all i . Thus, $L_{11} = \sqrt{P_{11}}$, as well as its reciprocal $1/L_{11}$, is smooth.

Now assume L_{ij} and $1/L_{jj}$ are smooth for $i = 1, \dots, i_0$ and $j = 1, \dots, j_0 \leq i_0$. As we just showed, this hypothesis is true for $i_0 = 1$ and $j_0 = 1$. If $j_0 = i_0$, from (2.1) we see that $L_{i_0+1,1} = (1/L_{11})P_{i_0+1,1}$ is smooth. If $j_0 < i_0 - 1$, then L_{i_0,j_0+1} results from a sequence of elementary operations, such as multiplication, addition, and subtraction, of maps $L_{i_0,1}, \dots, L_{i_0,j_0}, L_{j_0+1,1}, \dots, L_{j_0+1,j_0}$, and $1/L_{j_0+1,j_0+1}$ that are all smooth according to the induction hypothesis. As these elementary operations are all smooth, L_{i_0,j_0+1} is also smooth. If $j_0 = i_0 - 1$, then $L_{i_0,j_0+1} = L_{i_0,i_0}$, as well as $1/L_{i_0,i_0}$, is smooth via similar reasoning based on the additional fact that $P_{i_0,i_0} - \sum_{k=1}^{i_0-1} L_{ik}^2 > 0$ and the square-root operator $\sqrt{\cdot}$ is smooth on the set of positive real numbers. The above derivation then shows that the induction hypothesis is also true for $i = i_0, j = j_0 + 1$ if $j_0 < i_0$ and $i = i_0 + 1, j = 1$ if $j_0 = i_0$.

Consequently, by mathematical induction, the hypothesis is true for all pairs of i and $j \leq i$. In other words, \mathcal{L} is a smooth manifold map. Its inverse, denoted by \mathcal{S} , is given by $\mathcal{S}(L) = LL^\top$ and clearly smooth. Therefore, \mathcal{L} and its inverse \mathcal{S} are diffeomorphisms. \square

3. Lie group structure and bi-invariant metric. In this section, we first construct a group structure and a bi-invariant metric on the manifold \mathcal{L}_+ and then push them forward to the manifold \mathcal{S}_m^+ via the Cholesky map. Parallel transport on SPD manifolds is also investigated. For a background of Riemannian geometry and Lie group, we recommend monographs [19, 24].

3.1. Riemannian geometry on Cholesky spaces. For matrices in \mathcal{L}_+ , as off-diagonal elements in the lower triangular part are unconstrained while diagonal ones are restricted to be positive, \mathcal{L}_+ can be parameterized by $\mathcal{L} \ni X \mapsto \varphi(X) \in \mathcal{L}_+$ in the way that $(\varphi(X))_{ij} = X_{ij}$ if $i \neq j$ and $(\varphi(X))_{jj} = \exp(X_{jj})$. This motivates us to, respectively, endow the unconstrained part $\lfloor \cdot \rfloor$ and the positive diagonal part $\mathbb{D}(\cdot)$ with a different metric and then combine them into a Riemannian metric on \mathcal{L}_+ as follows. First, we note that the tangent space of \mathcal{L}_+ at a given $L \in \mathcal{L}_+$ is identified with the linear space \mathcal{L} . For such tangent space, we treat the strict lower triangular space $\lfloor \mathcal{L} \rfloor := \{ \lfloor X \rfloor : X \in \mathcal{L} \}$ as the Euclidean space $\mathbb{R}^{m(m-1)/2}$ with the usual Frobenius inner product $\langle X, Y \rangle_F = \sum_{i,j=1}^m X_{ij} Y_{ij}$ for all $X, Y \in \lfloor \mathcal{L} \rfloor$. For the diagonal part $\mathbb{D}(\mathcal{L}) := \{ \mathbb{D}(Z) : Z \in \mathcal{L} \}$, we equipped it with a different inner product defined by $\langle \mathbb{D}(L)^{-1}\mathbb{D}(X), \mathbb{D}(L)^{-1}\mathbb{D}(Y) \rangle_F$. Finally, combining these two components together, we define a metric \tilde{g} for tangent spaces $T_L \mathcal{L}_+$ (identified with \mathcal{L}) by

$$\begin{aligned} \tilde{g}_L(X, Y) &= \langle \lfloor X \rfloor, \lfloor Y \rfloor \rangle_F + \langle \mathbb{D}(L)^{-1}\mathbb{D}(X), \mathbb{D}(L)^{-1}\mathbb{D}(Y) \rangle_F \\ &= \sum_{i>j} X_{ij} Y_{ij} + \sum_{j=1}^m X_{jj} Y_{jj} L_{jj}^{-2}. \end{aligned}$$

It is straightforward to show that the space \mathcal{L}_+ , equipped with the metric \tilde{g} , is a Riemannian manifold. We begin with geodesics on the manifold $(\mathcal{L}_+, \tilde{g})$ to investigate its basic properties.

PROPOSITION 3. *On the Riemannian manifold $(\mathcal{L}_+, \tilde{g})$, the geodesic starting at $L \in \mathcal{L}_+$ with direction $X \in T_L \mathcal{L}_+$ is given by*

$$\tilde{\gamma}_{L,X}(t) = [L] + t[X] + \mathbb{D}(L) \exp\{t\mathbb{D}(X)\mathbb{D}(L)^{-1}\}.$$

Proof. Clearly, $\tilde{\gamma}_{L,X}(0) = L$ and $\tilde{\gamma}'_{L,X}(0) = [X] + \mathbb{D}(X) = X$. Now, we use $\text{vec}(L)$ to denote the vector in $\mathbb{R}^{m(m+1)/2}$ such that the first m elements of $\text{vec}(L)$ correspond to the diagonal elements of L . Define the map $x : \mathcal{L}_+ \rightarrow \mathbb{R}$ by

$$x^i(L) = \begin{cases} \log \text{vec}(L)_i & \text{if } 1 \leq i \leq m, \\ \text{vec}(L)_i, & \text{otherwise,} \end{cases}$$

where x^i denotes the i th component of x . It can be checked that (\mathcal{L}_+, x) is a chart for the manifold \mathcal{L}_+ . Let e_i be the $m(m+1)/2$ dimensional vector whose i th element is one and other elements are all zero. For $1 \leq i \leq m$, we define $\partial_i = \text{vec}(L)_i e_i$, and for $i > m$, define $\partial_i = e_i$. The collection $\{\partial_1, \dots, \partial_{m(m+1)/2}\}$ is a frame. One can check that $\tilde{g}_{ij} := \tilde{g}_L(\partial_i, \partial_j) = 0$ if $i \neq j$, and $\tilde{g}_{ii} = 1$. This implies that $\partial \tilde{g}_{jk}/\partial x^l = 0$, and hence all Christoffel symbols are zero, as

$$\Gamma^i_{kl} = \frac{1}{2} \tilde{g}^{ij} \left(\frac{\partial \tilde{g}_{jk}}{\partial x^l} + \frac{\partial \tilde{g}_{jl}}{\partial x^k} - \frac{\partial \tilde{g}_{kl}}{\partial x^j} \right) = 0,$$

where Einstein summation convention is assumed. It can be checked that the i th coordinate $\tilde{\gamma}^i(t) = x^i \circ \tilde{\gamma}_{L,X}(t)$ of the curve $\tilde{\gamma}_{L,X}$ is given by $\tilde{\gamma}^i(t) = \log \text{vec}(L)_i + t \text{vec}(X)_i / \text{vec}(L)_i$ when $i \leq m$ and $\tilde{\gamma}^i(t) = \text{vec}(L)_i + t \text{vec}(X)_i$ if $i > m$. Now, it is an easy task to verify the following geodesic equations

$$\frac{d^2 \tilde{\gamma}^i}{dt^2} + \Gamma^i_{jk} \frac{d\tilde{\gamma}^j}{dt} \frac{d\tilde{\gamma}^k}{dt} = 0$$

for $i = 1, \dots, m(m+1)/2$. Therefore, $\tilde{\gamma}_{L,X}(t)$ is the claimed geodesic. \square

Given the above proposition, we can immediately derive the Riemannian exponential map $\widetilde{\text{Exp}}$ at $L \in \mathcal{L}_+$, which is given by

$$\widetilde{\text{Exp}}_L X = \tilde{\gamma}_{L,X}(1) = [L] + [X] + \mathbb{D}(L) \exp\{\mathbb{D}(X)\mathbb{D}(L)^{-1}\}.$$

Also, for $L, K \in \mathcal{L}_+$, with $X = [K] - [L] + \{\log \mathbb{D}(K) - \log \mathbb{D}(L)\}\mathbb{D}(L)$, one has

$$\tilde{\gamma}_{L,X}(t) = [L] + t\{[K] - [L]\} + \mathbb{D}(L) \exp[t\{\log \mathbb{D}(K) - \log \mathbb{D}(L)\}].$$

Since $\tilde{\gamma}_{L,X}(0) = L$ and $\tilde{\gamma}_{L,X}(1) = K$, $\tilde{\gamma}_{L,X}$ is the geodesic connecting L and K . Therefore, the distance function on \mathcal{L}_+ induced by \tilde{g} , denoted by $d_{\mathcal{L}_+}$, is given by

$$d_{\mathcal{L}_+}(L, K) = \{\tilde{g}_L(X, X)\}^{1/2} = \left\{ \sum_{i>j} (L_{ij} - K_{ij})^2 + \sum_{j=1}^m (\log L_{jj} - \log K_{jj})^2 \right\}^{1/2},$$

where X is the same as the above. The expression for the distance function can be equivalently and more compactly written as

$$d_{\mathcal{L}_+}(L, K) = \{\|[L] - [K]\|_{\text{F}}^2 + \|\log \mathbb{D}(L) - \log \mathbb{D}(K)\|_{\text{F}}^2\}^{1/2}.$$

Table 1 summarizes the above basic properties of the manifold $(\mathcal{L}_+, \tilde{g})$.

TABLE 1
Basic properties of Riemannian manifolds $(\mathcal{L}_+, \tilde{g})$.

tangent space at L \mathcal{L}
Riemannian metric $\tilde{g}_L(X, Y) = \sum_{i>j} X_{ij}Y_{ij} + \sum_{j=1}^m X_{jj}Y_{jj}L_{jj}^{-2}$
geodesic emanating from L with direction X $\tilde{\gamma}_{L,X}(t) = [L] + t[X] + \mathbb{D}(L) \exp\{t\mathbb{D}(X)\mathbb{D}(L)^{-1}\}$
Riemannian exponential map at L $\widetilde{\text{Exp}}_L X = [L] + [X] + \mathbb{D}(L) \exp\{\mathbb{D}(X)\mathbb{D}(L)^{-1}\}$
Riemannian logarithmic map at L $\widetilde{\text{Log}}_L K = [K] - [L] + \mathbb{D}(L) \log\{\mathbb{D}(L)^{-1}\mathbb{D}(K)\}$
geodesic distance between L and K $d_{\mathcal{L}_+}(L, K) = \{\ [L] - [K]\ _{\mathbb{F}}^2 + \ \log \mathbb{D}(L) - \log \mathbb{D}(K)\ _{\mathbb{F}}^2\}^{1/2}$

3.2. Riemannian metric for SPD matrices. As mentioned previously, the space \mathcal{S}_m^+ of SPD matrices is a smooth submanifold of the space \mathcal{S}_m of symmetric matrices, whose tangent space at a given SPD matrix is identified with \mathcal{S}_m . We also showed that the map $\mathcal{S} : \mathcal{L}_+ \rightarrow \mathcal{S}_m^+$ by $\mathcal{S}(L) = LL^\top$ is a diffeomorphism between \mathcal{L}_+ and \mathcal{S}_m^+ . For a square matrix S , we define a lower triangular matrix $S_{\frac{1}{2}} = [S] + \mathbb{D}(S)/2$. In other words, the matrix $S_{\frac{1}{2}}$ is the lower triangular part of S with the diagonal elements halved. The differential of \mathcal{S} is given in the following.

PROPOSITION 4. *The differential $D_L \mathcal{S} : T_L \mathcal{L}_+ \rightarrow T_{LL^\top} \mathcal{S}_m^+$ of \mathcal{S} at L is given by*

$$(D_L \mathcal{S})(X) = LX^\top + XL^\top.$$

Also, the inverse $(D_L \mathcal{S})^{-1} : T_{LL^\top} \mathcal{S}_m^+ \rightarrow T_L \mathcal{L}_+$ of $D_L \mathcal{S}$ exists for all $L \in \mathcal{L}_+$ and is given by

$$(D_L \mathcal{S})^{-1}(W) = L(L^{-1}WL^{-\top})_{\frac{1}{2}}$$

for $W \in \mathcal{S}_m$.

Proof. Let $X \in \mathcal{L}$ and $L \in \mathcal{L}_+$. Then $\tilde{\gamma}_L(t) = L + tX$ is a curve passing through L if $t \in (-\epsilon, \epsilon)$ for a sufficiently small $\epsilon > 0$. Note that for every such t , $\tilde{\gamma}_L(t) \in \mathcal{L}_+$. Then $\gamma_{LL^\top}(t) = \mathcal{S}(\tilde{\gamma}_L(t))$ is a curve passing through LL^\top . The differential is then derived from

$$(D_L \mathcal{S})(X) = (\mathcal{S} \circ \tilde{\gamma}_L)'(0) = \frac{d}{dt} \mathcal{S}(\tilde{\gamma}_L(t))|_{t=0} = LX^\top + XL^\top.$$

On the other hand, if $W = LX^\top + XL^\top$, then since L is invertible, we have

$$L^{-1}WL^{-\top} = X^\top L^{-\top} + L^{-1}X = L^{-1}X + (L^{-1}X)^\top.$$

Note that $L^{-1}X$ is also a lower triangular matrix, and the matrix on the left-hand side is symmetric, we deduce that $(L^{-1}WL^{-\top})_{\frac{1}{2}} = L^{-1}X$, which gives rise to $X = L(L^{-1}WL^{-\top})_{\frac{1}{2}}$. The linear map $(D_L \mathcal{S})(X) = LX^\top + XL^\top$ is one-to-one, since from the above derivation, $(D_L \mathcal{S})(X) = 0$ if and only if $X = 0$. \square

Given the above proposition, the manifold map \mathcal{S} , which is exactly the inverse of the Cholesky map \mathcal{L} discussed in subsection 2.2, induces a Riemannian metric on \mathcal{S}_m^+ , denoted by g and called *Log-Cholesky metric*, given by

$$(3.1) \quad g_P(W, V) = \tilde{g}_L \left(L(L^{-1}WL^{-\top})_{\frac{1}{2}}, L(L^{-1}VL^{-\top})_{\frac{1}{2}} \right),$$

where $L = \mathcal{S}^{-1}(P) = \mathcal{L}(P)$ is the Cholesky factor of $P \in \mathcal{S}_m^+$, and $W, V \in \mathcal{S}_m$ are tangent vectors at P . This implies that

$$\tilde{g}_L(X, Y) = g_{\mathcal{S}(L)}((D_L\mathcal{S})(X), (D_L\mathcal{S})(Y))$$

for all L and $X, Y \in T_L\mathcal{L}_+$. According to Definition 7.57 of [26], the map \mathcal{S} is an isometry between $(\mathcal{L}_+, \tilde{g})$ and (\mathcal{S}_m^+, g) . A Riemannian isometry provides correspondence of Riemannian properties and objects between two Riemannian manifolds. This enables us to study the properties of (\mathcal{S}_m^+, g) via the manifold $(\mathcal{L}_+, \tilde{g})$ and the isometry \mathcal{S} . For example, we can obtain geodesics on \mathcal{S}_m^+ by mapping geodesics on \mathcal{L}_+ . More precisely, the geodesic emanating from $P = LL^\top$ with $L = \mathcal{L}(P)$ is given by

$$\gamma_{P,W}(t) = \mathcal{S}(\tilde{\gamma}_{L,X}(t)) = \tilde{\gamma}_{L,X}(t)\tilde{\gamma}_{L,X}^\top(t),$$

where $X = L(L^{-1}WL^{-\top})_{\frac{1}{2}}$ and $W \in T_P\mathcal{S}_m^+$. Similarly, the Riemannian exponential at P is given by

$$\text{Exp}_P W = \mathcal{S}(\widetilde{\text{Exp}}_L X) = (\widetilde{\text{Exp}}_L X)(\widetilde{\text{Exp}}_L X)^\top,$$

while the geodesic between P and Q is characterized by

$$\gamma_{P,W}(t) = \tilde{\gamma}_{L,X}(t)\tilde{\gamma}_{L,X}^\top(t),$$

with $L = \mathcal{L}(P)$, $K = \mathcal{L}(Q)$, $X = \lfloor K \rfloor - \lfloor L \rfloor + \{\log \mathbb{D}(K) - \log \mathbb{D}(L)\}\mathbb{D}(L)$, and $W = LX^\top + XL^\top$. Also, the geodesic distance between P and Q is

$$d_{\mathcal{S}_m^+}(P, Q) = d_{\mathcal{L}_+}(\mathcal{L}(P), \mathcal{L}(Q)).$$

Moreover, the Levi-Civita connection ∇ of (\mathcal{S}_m^+, g) can be obtained by the Levi-Civita connection $\tilde{\nabla}$ of $(\mathcal{L}_+, \tilde{g})$. To see this, let W and V be two smooth vector fields on \mathcal{S}_m^+ . Define vector fields X and Y on \mathcal{L}_+ by $X(L) = (D_{LL^\top}\mathcal{L})W(LL^\top)$ and $Y(L) = (D_{LL^\top}\mathcal{L})V(LL^\top)$ for all $L \in \mathcal{L}_+$. Then $\nabla_W V = (D\mathcal{S})(\tilde{\nabla}_X Y)$, and the Christoffel symbols to compute the connection $\tilde{\nabla}$ has been given in the proof of Proposition 3.

Table 2 summarizes some basic properties of the manifold (\mathcal{S}_m^+, g) . Note that the differential $D_P\mathcal{L}$ can be computed efficiently, since it only involves Cholesky decomposition and the inverse of a lower triangular matrix, for both of which there exist efficient algorithms. Consequently, all maps in Table 2 can be evaluated in an efficient way. In contrast, computation of Riemannian exponential/logarithmic maps for the Log-Euclidean metric [4] requires evaluation of some series of an infinite number of terms; see (2.1) and Table 4.1 of [4].

3.3. Lie group structure and bi-invariant metrics . We define an operator \odot on \mathcal{L} by

$$X \odot Y = \lfloor X \rfloor + \lfloor Y \rfloor + \mathbb{D}(X)\mathbb{D}(Y).$$

Note that $\mathcal{L}_+ \subset \mathcal{L}$. Moreover, if $L, K \in \mathcal{L}_+$, then $L \odot K \in \mathcal{L}_+$. It is not difficult to see that \odot is a smooth commutative group operation on the manifold \mathcal{L}_+ , where the

TABLE 2
Properties of Riemannian manifold (\mathcal{S}_m^+, g) .

tangent space at P \mathcal{S}_m
differential of \mathcal{S} at L $D_L \mathcal{S} : X \mapsto LX^\top + XL^\top$
differential of \mathcal{L} at P $D_P \mathcal{L} : W \mapsto \mathcal{L}(P)(\mathcal{L}(P)^{-1}W\mathcal{L}(P)^{-\top})_{\frac{1}{2}}$
Riemannian metric $g_P(W, V) = \tilde{g}_{\mathcal{L}(P)}((D_P \mathcal{L})(W), (D_P \mathcal{L})(V))$
geodesic emanating from P with direction W $\gamma_{P,W}(t) = \tilde{\gamma}_{\mathcal{L}(P), (D_P \mathcal{L})(W)}(t)\tilde{\gamma}_{\mathcal{L}(P), (D_P \mathcal{L})(W)}^\top(t)$
Riemannian exponential map at P $\text{Exp}_P W = \widetilde{\text{Exp}}_{\mathcal{L}(P)}(D_P \mathcal{L})(W)\{\widetilde{\text{Exp}}_{\mathcal{L}(P)}(D_P \mathcal{L})(W)\}^\top$
Riemannian logarithmic map at P $\text{Log}_P Q = (D_{\mathcal{L}(P)} \mathcal{S})(\widetilde{\text{Log}}_{\mathcal{L}(P)} \mathcal{L}(Q))$
geodesic distance between P and Q $d_{\mathcal{S}_m^+}(P, Q) = d_{\mathcal{L}_+}(\mathcal{L}(P), \mathcal{L}(Q)).$

inverse of L , denoted by L_\otimes^{-1} , is $\mathbb{D}(L)^{-1} - \lfloor L \rfloor$. The left translation by $A \in \mathcal{L}_+$ is denoted by $\ell_A : B \mapsto A \otimes B$. One can check that the differential of this operation at $L \in \mathcal{L}_+$ is

$$(3.2) \quad D_L \ell_A : X \mapsto \lfloor X \rfloor + \mathbb{D}(A)\mathbb{D}(X),$$

where it is noted that the differential $D_L \ell_A$ does not depend on L . Given the above expression, one can find that

$$\begin{aligned} & \tilde{g}_{A \otimes L}((D_L \ell_A)(X), (D_L \ell_A)(Y)) \\ &= \tilde{g}_{A \otimes L}(\lfloor X \rfloor + \mathbb{D}(A)\mathbb{D}(X), \lfloor Y \rfloor + \mathbb{D}(A)\mathbb{D}(Y)) \\ &= \tilde{g}_L(X, Y). \end{aligned}$$

Similar observations are made for right translations. Thus, the metric \tilde{g} is a bi-invariant metric that turns (\mathcal{L}_+, \otimes) into a Lie group.

The group operator \otimes and maps \mathcal{S} and \mathcal{L} together induce a smooth operation \circledast on \mathcal{S}_m^+ , defined by

$$\begin{aligned} P \circledast Q &= \mathcal{S}(\mathcal{L}(P) \circledast \mathcal{L}(Q)) \\ &= (\mathcal{L}(P) \circledast \mathcal{L}(Q))(\mathcal{L}(P) \circledast \mathcal{L}(Q))^\top \quad \text{for } P, Q \in \mathcal{S}_m^+. \end{aligned}$$

In addition, both \mathcal{L} and \mathcal{S} are Riemannian isometries and group isomorphisms between Lie groups $(\mathcal{L}_+, \tilde{g}, \otimes)$ and $(\mathcal{S}_m^+, g, \circledast)$.

THEOREM 5. *The space $(\mathcal{S}_m^+, \circledast)$ is an abelian Lie group. Moreover, the metric g defined in (3.1) is a bi-invariant metric on $(\mathcal{S}_m^+, \circledast)$.*

Proof. It is clear that \mathcal{S}_m^+ is closed under the operation \circledast , and the identity element is the identity matrix. For $P \in \mathcal{S}_m^+$, the inverse under \circledast is given by $(\mathcal{L}(P)_{\circledast}^{-1})(\mathcal{L}(P)_{\circledast}^{-1})^\top$. For associativity, we first observe that $\mathcal{L}(P \circledast Q) = \mathcal{L}(P) \circledast \mathcal{L}(Q)$, based on which we further deduce that

$$\begin{aligned} (P \circledast Q) \circledast S &= (\mathcal{L}(P \circledast Q) \circledast \mathcal{L}(S))(\mathcal{L}(P \circledast Q) \circledast \mathcal{L}(S))^\top \\ &= (\mathcal{L}(P) \circledast \mathcal{L}(Q) \circledast \mathcal{L}(S))(\mathcal{L}(P) \circledast \mathcal{L}(Q) \circledast \mathcal{L}(S))^\top \\ &= (\mathcal{L}(P) \circledast \mathcal{L}(Q \circledast S))(\mathcal{L}(P) \circledast \mathcal{L}(Q \circledast S))^\top \\ &= P \circledast (Q \circledast S). \end{aligned}$$

Therefore, $(\mathcal{S}_m^+, \circledast)$ is a group. The commutativity and smoothness of \circledast stem from the commutativity and smoothness of \circledast , respectively. It can be checked that \mathcal{S} is a group isomorphism and isometry between Lie groups $(\mathcal{L}_+, \circledast)$ and $(\mathcal{S}_m^+, \circledast)$, respectively, endowed with Riemannian metrics \tilde{g} and g . Then, the bi-invariance of g follows from the bi-invariance of \tilde{g} . \square

3.4. Parallel transport along geodesics on \mathcal{S}_m^+ . In some applications like statistical analysis or machine learning on Riemannian manifolds, parallel transport of tangent vectors along geodesics is required. For instance, in [40] that studies regression on SPD-valued data, tangent vectors are transported to a common place to derive statistical estimators of interest. Also, optimization in the context of statistics for manifold-valued data often involves parallel transport of tangent vectors. Examples include Hamiltonian Monte Carlo algorithms [23] as well as optimization algorithms [20] to train manifold-valued Gaussian mixture models. In these scenarios, Riemannian metrics on SPD matrices that result in efficient computation for parallel transport are attractive, in particular in the era of big data. In this regard, as discussed in the introduction, evaluation of parallel transport along geodesics for the affine-invariant metric is simple and efficient in computation, while the one for the Log-Euclidean metric is computationally intensive. Below we show that parallel transport for the presented metric also has a simple form, starting with a lemma.

LEMMA 6. *Let (\mathcal{G}, \cdot) be an abelian Lie group with a bi-invariant metric. The parallel transporter $\tau_{p,q}$ that transports tangent vectors at p to tangent vectors at q along the geodesic connecting p and q is given by $\tau_{p,q}(u) = (D_p \ell_{q,p^{-1}})u$ for $u \in T_p \mathcal{G}$.*

Proof. For simplicity, we abbreviate $p \cdot q$ as pq . Let \mathfrak{g} denote the Lie algebra associated with the Lie group \mathcal{G} , and ∇ the Levi–Civita connection on \mathcal{G} . Note that we identify elements in \mathfrak{g} with left-invariant vector fields on \mathcal{G} . We shall first recall that $\nabla_Y Z = [Y, Z]/2$ for $Y, Z \in \mathfrak{g}$ (see the proof of Theorem 21.3 in [29]), where $[\cdot, \cdot]$ denotes the Lie bracket of \mathcal{G} . As \mathcal{G} is abelian, the Lie bracket vanishes everywhere, and hence $\nabla_Y Z = 0$ if $Y, Z \in \mathfrak{g}$.

Let $\gamma_p(t) = \ell_p(\exp(tY))$ for $Y \in \mathfrak{g}$ such that $\exp(Y) = p^{-1}q$, where \exp denotes the Lie group exponential map. Recall that for a bi-invariant Lie group, the group exponential map coincides with the Riemannian exponential map at the group identity e , and left translations are isometries. Thus, γ_p is a geodesic. Using the fact $\gamma_e(t+s) = \gamma_e(t)\gamma_e(s) = \ell_{\gamma_e(t)}(\gamma_e(s))$ according to Lemma 21.2 of [29], by the chain rule of differential, we have

$$\gamma'_e(t) = \frac{d}{dt} \gamma_e(t+s) |_{s=0} = (D_e \ell_{\gamma_e(t)}) (\gamma'_e(0)) = (D_e \ell_{\gamma_e(t)}) (Y(e)) = Y(\gamma_e(t)),$$

from which we further deduce that

$$\gamma'_p(t) = (D_{\gamma_e(t)} \ell_p) \gamma'_e(t) = (D_{\gamma_e(t)} \ell_p) Y(\gamma_e(t)) = Y(p\gamma_e(t)).$$

Now define a vector field $Z(q) := (D_p \ell_{qp^{-1}})u$. We claim that Z is a left-invariant vector field on \mathcal{G} and hence belongs to \mathfrak{g} , since

$$\begin{aligned} Z(hq) &= (D_p \ell_{hqp^{-1}})u = \{D_p(\ell_h \circ \ell_{qp^{-1}})\}u \\ &= (D_q \ell_h)(D_p \ell_{qp^{-1}})u = (D_q \ell_h)(Z(q)), \end{aligned}$$

where the third equality is obtained by the chain rule of differential. Consequently, $\nabla_{\gamma'_p} Z = 0$ since $\gamma'_p(t) = Y(p\gamma_e(t))$ and $\nabla_Y Z = 0$ for $Y, Z \in \mathfrak{g}$. As additionally $Z(\gamma_p(0)) = Z(p) = u$, transportation of u along the geodesic γ_p is realized by the left-invariant vector field Z . Since γ_p is a geodesic with $\gamma_p(0) = p$ and $\gamma_p(1) = \ell_p \exp(Y) = p(p^{-1}q) = q$, we have that

$$\tau_{p,q}(u) = Z(\gamma_p(1)) = Z(q) = (D_p \ell_{qp^{-1}})u,$$

as claimed. \square

PROPOSITION 7. *A tangent vector $X \in T_L \mathcal{L}_+$ is parallelly transported to the tangent vector $\lfloor X \rfloor + \mathbb{D}(K)\mathbb{D}(L)^{-1}\mathbb{D}(X)$ at K . For $P, Q \in \mathcal{S}_m^+$ and $W \in \mathcal{S}_m$,*

$$\tau_{P,Q}(W) = K\{\lfloor X \rfloor + \mathbb{D}(K)\mathbb{D}(L)^{-1}\mathbb{D}(X)\}^\top + \{\lfloor X \rfloor + \mathbb{D}(K)\mathbb{D}(L^{-1})\mathbb{D}(X)\}K^\top,$$

where $L = \mathcal{L}(P)$, $K = \mathcal{L}(Q)$ and $X = L(L^{-1}WL^{-\top})_{\frac{1}{2}}$.

Proof. By Lemma 6, it is seen that $\tau_{L,K}(X) = (D_L \ell_{K \circledast L^{-1}})X$. According to (3.2),

$$(D_L \ell_{K \circledast L^{-1}})X = \lfloor X \rfloor + \mathbb{D}(K \circledast L^{-1})\mathbb{D}(X) = \lfloor X \rfloor + \mathbb{D}(K)\mathbb{D}(L)^{-1}\mathbb{D}(X).$$

The statement for P, Q, W follows from the fact that \mathcal{S} and \mathcal{L} are an isometries. \square

The above proposition shows that parallel transport for the presented Log-Cholesky metric can be computed rather efficiently. In fact, the only nontrivial steps in computation are to perform Cholesky decomposition of two SPD matrices and to invert a lower triangular matrix, for both of which there exist efficient algorithms. It is numerically faster than the affine-invariant metric and Log-Euclidean metric. For instance, for 5×5 SPD matrices, in a MATLAB computational environment, on average it takes 9.3 ms (Log-Euclidean), 0.85 ms (affine-invariant) and 0.2 ms (Log-Cholesky) on an Intel(R) Core i7-4500U (1.80 GHz) to perform a parallel transport. We see that, to do parallel transport, the Log-Euclidean metric is about 45 times slower than the Log-Cholesky metric, since it has to evaluate the differential of a matrix logarithmic map that is expressed as a convergent series of infinite terms of products of matrices. In practice, only a finite leading terms are evaluated. However, in order to avoid significant loss of precision, a large number of terms are often needed. For instance, for a precision of the order of 10^{-12} , typically approximately 300 leading terms are required. The affine-invariant metric, despite having a simple form for parallel transport, is still about 4 times slower than our Log-Cholesky metric, partially due to that more matrix inversion and multiplication operations are needed.

4. Mean of distributions on \mathcal{S}_m^+ . In this section we study the Log-Cholesky mean of a random SPD matrix and the Log-Cholesky average of a finite number of SPD matrices. We first establish the existence and uniqueness of such quantities. A closed and easy-to-compute form for Log-Cholesky averages is also derived. Finally, we show that determinants of Log-Cholesky averages are bounded by determinants of SPD matrices being averaged. This property suggests that Log-Cholesky averages are not subject to swelling effect.

4.1. Log-Cholesky mean of a random SPD matrix. For a random element Q on a Riemannian manifold \mathcal{M} , we define a function $F(x) = \mathbb{E}d_{\mathcal{M}}^2(x, Q)$, where $d_{\mathcal{M}}$ denotes the geodesic distance function on \mathcal{M} , and \mathbb{E} denotes expectation of a random number. If $F(x) < \infty$ for some $x \in \mathcal{M}$,

$$x = \arg \min_{z \in \mathcal{M}} F(z),$$

and $F(z) \geq F(x)$ for all $z \in \mathcal{M}$, then x is called a Fréchet mean of Q , denoted by $\mathbb{E}Q$. In general, Fréchet mean might not exist, and when it exists, it might not be unique; see [1] for conditions of the existence and uniqueness of Fréchet means. However, for \mathcal{S}_m^+ endowed with Log-Cholesky metric, we claim that if $F(S) < \infty$ for some $S \in \mathcal{S}_m^+$, then the Fréchet mean exists and is unique. Such Fréchet mean is termed *Log-Cholesky mean* in this paper. To prove the claim, we first notice that, similar to the Log-Euclidean metric [4], the manifold (\mathcal{S}_m^+, g) is a “flat” space.

PROPOSITION 8. *The sectional curvature of $(\mathcal{L}_+, \tilde{g})$ and (\mathcal{S}_m^+, g) is constantly zero.*

Proof. For $(\mathcal{L}_+, \tilde{g})$, in the proof of Proposition 3, it has been shown that all Christoffel symbols are zero under the selected coordinate. This implies that the Riemannian curvature tensor is identically zero and hence so is the sectional curvature. The case of (\mathcal{S}_m^+, g) follows from the fact that \mathcal{S} is an isometry that preserves sectional curvature. \square

PROPOSITION 9. *If L is a random element on $(\mathcal{L}_+, \tilde{g})$ and $\mathbb{E}d_{\mathcal{L}_+}^2(A, L) < \infty$ for some $A \in \mathcal{L}_+$. Then the Fréchet mean $\mathbb{E}L$ exists and is unique. Similarly, the Log-Cholesky mean of a random SPD matrix P exists and is unique if $\mathbb{E}d_{\mathcal{S}_m^+}^2(S, P) < \infty$ for some SPD matrix S .*

Proof. In the case of $(\mathcal{L}_+, \tilde{g})$, we define $\psi(L) = \lfloor L \rfloor + \log \mathbb{D}(L)$. It can be shown that ψ is a diffeomorphism (and hence also a homeomorphism) between \mathcal{L}_+ and \mathcal{L} . Therefore, \mathcal{L}_+ is simply connected since \mathcal{L} is. In Proposition 8 we have shown that \mathcal{L}_+ has zero sectional curvature. Thus, the existence and uniqueness of Fréchet mean follows from Theorem 2.1 of [8]. The case of (\mathcal{S}_m^+, g) follows from the isometry of \mathcal{S} . \square

The next result shows that the Log-Cholesky mean of a random SPD matrix is computable from the Fréchet mean of its Cholesky factor. Also, it characterizes the Fréchet mean of a random Cholesky factor, which is important for us to derive a closed form for the Log-Cholesky average of a finite number of matrices in the next subsection.

PROPOSITION 10. *Suppose L is a random element on \mathcal{L}_+ and P is a random element on \mathcal{S}_m^+ . Suppose for some fixed $A \in \mathcal{L}_+$ and $B \in \mathcal{S}_m^+$ such that $\mathbb{E}d_{\mathcal{L}_+}^2(A, L) < \infty$ and $\mathbb{E}d_{\mathcal{S}_m^+}^2(B, P) < \infty$. Then the Fréchet mean of L is given by*

$$(4.1) \quad \mathbb{E}L = \mathbb{E}\lfloor L \rfloor + \exp\{\mathbb{E}\log \mathbb{D}(L)\},$$

and the Log-Cholesky mean of P is given by

$$(4.2) \quad \mathbb{E}P = \{\mathbb{E}\mathcal{L}(P)\}\{\mathbb{E}\mathcal{L}(P)\}^\top.$$

Proof. Let $F(R) = \mathbb{E}d_{\mathcal{L}_+}^2(R, L)$. Then

$$F(R) = \mathbb{E}\|\lfloor R \rfloor - \lfloor L \rfloor\|_F^2 + \mathbb{E}\|\log \mathbb{D}(R) - \log \mathbb{D}(L)\|_F^2 := F_1(\lfloor R \rfloor) + F_2(\mathbb{D}(R)).$$

To minimize F , we can separately minimize F_1 and F_2 , due to two reasons: (1) F_1 involves $[R]$ only while F_2 involves $\mathbb{D}(R)$, and (2) $[R]$ and $\mathbb{D}(R)$ are disjoint and can vary independently. For F_1 , we first note that the condition $\mathbb{E}d_{\mathcal{L}_+}^2(A, L) < \infty$ ensures that $\mathbb{E}\|L\|_F^2 < \infty$, and hence the mean $\mathbb{E}[L]$ is well defined. Also, it can be checked that this mean minimizes F_1 .

For F_2 , we note that

$$\begin{aligned} F_2(\mathbb{D}(R)) &= \sum_{j=1}^m \mathbb{E}(\log \mathbb{D}(R) - \log \mathbb{D}(L))_{jj}^2 \\ &= \sum_{j=1}^m \mathbb{E}\{\log R_{jj} - \log L_{jj}\}^2 := \sum_{j=1}^m F_{2j}(R_{jj}). \end{aligned}$$

Again, these m components F_{2j} can be optimized separately. From the condition $\mathbb{E}d_{\mathcal{L}_+}^2(A, L) < \infty$ we deduce that $\mathbb{E}(\log L_{jj})^2 < \infty$. Thus, $\mathbb{E}\log L_{jj}$ is well defined and minimizes $F_{2j}(e^x)$ with respect to x . Therefore, $\exp\{\mathbb{E}\log L_{jj}\}$ minimizes $F_{2j}(x)$ for each $j = 1, \dots, m$. In matrix form, this is equivalent to that $\exp\{\mathbb{E}\log \mathbb{D}(L)\}$ minimizes $F_2(D)$ when D is constrained to be diagonal. Thus, combined with the optimizer for F_1 , it establishes (4.1). Then (4.2) follows from the fact that $\mathcal{S}(L) = LL^\top$ is an isometry. \square

Finally, we establish the following useful relation between the determinant of the Log-Cholesky mean and the mean of the logarithmic determinant of a random SPD matrix.

PROPOSITION 11. *If the Fréchet mean of a random element Q on $(\mathcal{L}_+, \tilde{g})$ or (\mathcal{S}_m^+, g) exists, then $\log \det(\mathbb{E}Q) = \mathbb{E}\log(\det Q)$.*

Proof. For the case that Q is a random element on $(\mathcal{L}_+, \tilde{g})$, we denote it by L and observe that in (4.1) $\mathbb{E}L$ is a lower triangular matrix and $\exp\{\mathbb{E}\log \mathbb{D}(L)\}$ is its diagonal part. For a triangular matrix, its determinant is the product of diagonal elements. Thus,

$$\det L = \prod_{j=1}^m L_{jj} = \exp[\text{tr}\{\log \mathbb{D}(L)\}],$$

or equivalently, $\log \det L = \text{tr}\{\log \mathbb{D}(L)\}$. The above observations imply that

$$\begin{aligned} \det \mathbb{E}L &= \exp[\text{tr}\{\log \mathbb{D}(\mathbb{E}L)\}] = \exp[\text{tr}\{\mathbb{E}\log \mathbb{D}(L)\}] \\ &= \exp[\mathbb{E}\text{tr}\{\log \mathbb{D}(L)\}] = \exp\{\mathbb{E}\log(\det L)\}, \end{aligned}$$

which proves the case of $(\mathcal{L}_+, \tilde{g})$.

For the case that Q is a random element on (\mathcal{S}_m^+, g) , we denote it by P and observe that $\det(AB) = (\det A)(\det B)$ for any square matrices A and B . Let $L = \mathcal{L}(P)$. Then, one has

$$\begin{aligned} \det \mathbb{E}P &= \det\{(\mathbb{E}L)(\mathbb{E}L)^\top\} \\ &= \{\det(\mathbb{E}L)\}\{\det(\mathbb{E}L)^\top\} \\ &= \exp\{\mathbb{E}\log(\det L)\}\exp\{\mathbb{E}\log(\det L^\top)\} \\ &= \exp\{\mathbb{E}(\log \det L + \log \det L^\top)\} \\ &= \exp[\mathbb{E}\log\{(\det L)\det(L^\top)\}], \end{aligned}$$

$$\begin{aligned} &= \exp \{ \mathbb{E} \log \det(LL^\top) \} \\ &= \exp \{ \mathbb{E} \log \det(P) \}, \end{aligned}$$

which establishes the statement for the case of (\mathcal{S}_m^+, g) . \square

4.2. Log-Cholesky average of finite SPD matrices. Let Q_1, \dots, Q_n be n points on a Riemannian manifold \mathcal{M} . The Fréchet average of these points, denoted by $\mathbb{E}_n(Q_1, \dots, Q_n)$, is defined to be the minimizer of function $F_n(x) = \sum_{i=1}^n d_{\mathcal{M}}^2(x, Q_i)$ if such minimizer exists and is unique. Clearly, this concept is analogous to the Fréchet mean of a random element discussed above. In fact, the set of the elements Q_1, \dots, Q_n , always corresponds to a random element Q with the uniform distribution on the set $\{Q_1, \dots, Q_n\}$. With this correspondence, the Fréchet average of Q_1, \dots, Q_n is simply the Fréchet mean of Q , i.e., $\mathbb{E}_n(Q_1, \dots, Q_n) = \mathbb{E}Q$. The following result is a corollary of Proposition 10 in conjunction with the above correspondence.

COROLLARY 12. *For $L_1, \dots, L_n \in \mathcal{L}_+$, one has*

$$(4.3) \quad \mathbb{E}_n(L_1, \dots, L_n) = \frac{1}{n} \sum_{i=1}^n \lfloor L_i \rfloor + \exp \left\{ n^{-1} \sum_{i=1}^n \log \mathbb{D}(L_i) \right\},$$

and for $P_1, \dots, P_n \in \mathcal{S}_m^+$, one has

$$(4.4) \quad \mathbb{E}_n(P_1, \dots, P_n) = \mathbb{E}_n\{\mathcal{L}(P_1), \dots, \mathcal{L}(P_n)\} \mathbb{E}_n\{\mathcal{L}(P_1), \dots, \mathcal{L}(P_n)\}^\top.$$

For a general manifold, Fréchet averages often do not admit a closed form. For example, no closed form of affine-invariant averages has been found. Strikingly, Log-Euclidean averages admit a simple and closed form which is attractive in applications. The above corollary shows that Log-Cholesky averages enjoy the same nice property of their Log-Euclidean counterparts. The following is a consequence of Proposition 11 and the aforementioned principle of correspondence between a set of objects and a random object with the discrete uniform probability distribution on them.

COROLLARY 13. *For $L_1, \dots, L_n \in \mathcal{L}_+$ and $P_1, \dots, P_n \in \mathcal{S}_m^+$, one has*

$$(4.5) \quad \det \mathbb{E}_n(L_1, \dots, L_n) = \exp \left(\frac{1}{n} \sum_{i=1}^n \log \det L_i \right)$$

and

$$(4.6) \quad \det \mathbb{E}_n(P_1, \dots, P_n) = \exp \left(\frac{1}{n} \sum_{i=1}^n \log \det P_i \right).$$

Equation (4.6) shows that the determinant of the Log-Cholesky average of n SPD matrices is the geometric mean of their determinants. Consequently, the Log-Cholesky average can be considered as a generalization of the geometric mean of SPD matrices [2], since according to [4], this property is “the common property that should have all generalizations of the geometric mean to SPD matrices.” Note that the Log-Cholesky average is also a Fréchet mean which is a generalization of the Euclidean mean and has applications in statistics on Riemannian manifolds [33], while the geometric mean [2] is algebraically constructed and not directly related to Riemannian geometry.

Corollary 13 also suggests that the determinant of the Log-Cholesky average is equal to the determinant of the Log-Euclidean and affine-invariant averages. Thus,

like these two averages, the Log-Cholesky average does not suffer from swelling effect. In fact, as a consequence of the above corollary, one has that

$$(4.7) \quad \inf_{1 \leq i \leq n} \det L_i \leq \det \mathbb{E}_n(L_1, \dots, L_n) \leq \sup_{1 \leq i \leq n} \det L_i$$

and

$$(4.8) \quad \inf_{1 \leq i \leq n} \det P_i \leq \det \mathbb{E}_n(P_1, \dots, P_n) \leq \sup_{1 \leq i \leq n} \det P_i.$$

To see so, we first observe that, for positive numbers x_1, \dots, x_n , one has

$$\inf_{1 \leq i \leq m} \log x_i \leq n^{-1} \sum_{i=1}^n \log x_i \leq \sup_{1 \leq i \leq m} \log x_i.$$

Then, the monotonicity of $\exp(x)$ implies that

$$\inf_{1 \leq i \leq m} x_i \leq \exp \left(n^{-1} \sum_{i=1}^n \log x_i \right) \leq \sup_{1 \leq i \leq m} x_i,$$

and both (4.7) and (4.8) follow from Corollary 13.

As argued in [4], proper interpolation of SPD matrices is of importance in diffusion tensor imaging. The analogy of linear interpolation for Riemannian manifolds is geodesic interpolation. Specifically, if P and Q are two points on a manifold and $\gamma(t)$ is a geodesic connecting them such that $\gamma(0) = P$ and $\gamma(1) = Q$, then we say that γ geodesically interpolates P and Q . This notion of linear interpolation via geodesics can be straightforwardly generalized to bilinear or higher dimensional linear interpolation of points on a manifold. In Figure 1, we present an illustration of such geodesic interpolation for SPD matrices under the Euclidean metric, Cholesky distance [11], affine-invariant metric, Log-Euclidean metric, and Log-Cholesky metric. The Euclidean case exhibits significant swelling effect. Comparing to the Euclidean case, the Cholesky distance [11] substantially alleviates the effect but still suffers from noticeable swelling effect. In contrast, the Log-Cholesky metric, as well as the affine-invariant metric and Log-Euclidean metric, is not subject to any swelling effect. In addition, the affine-invariant, Log-Euclidean, and Log-Cholesky geodesic interpolations showed in Figure 1 are visibly indifferent. In fact, numerical simulations confirm that these three metrics often yield a similar Fréchet average of SPD matrices. For example, for a set of 20 randomly generated SPD matrices of dimension $m = 3$, the expected relative difference in terms of squared Frobenius norm between the Log-Cholesky average and the affine-invariant(or Log-Euclidean) average is approximately 3.3×10^{-2} .

The computation of geodesic interpolation for the Log-Cholesky metric is as efficient as the one for the Log-Euclidean metric, since both metrics enjoy a simple closed form for the Fréchet average of finite SPD matrices. Moreover, it is even numerically stabler than the Log-Euclidean metric which is in turn stabler than the affine-invariant metric. On synthetic examples of 3×3 SPD matrices with the largest eigenvalue 10^{10} (resp., 10^{15}) times larger than the smallest eigenvalue, the Log-Cholesky metric is still stable, while the Log-Euclidean one starts to deteriorate (resp., numerically collapse).

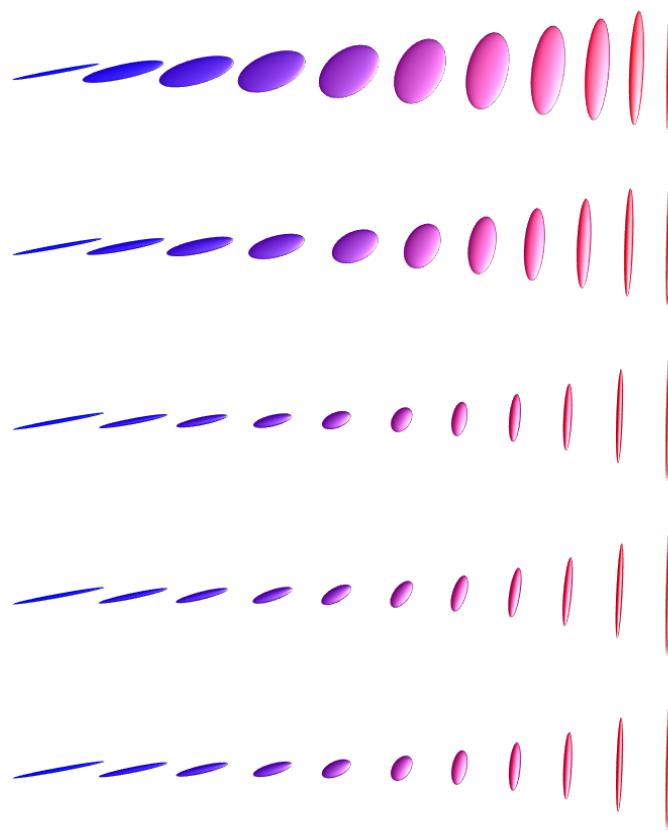


FIG. 1. *Interpolation of SPD matrices.* Top: Euclidean linear interpolation. The associated determinants are 5.40, 17.92, 27.68, 34.69, 38.93, 40.41, 39.14, 35.11, 28.32, 18.77, 6.46. Clearly, Euclidean interpolation exhibits significant swelling effect. Second row: Cholesky interpolation. 5.40, 9.31, 13.12, 16.29, 18.43, 19.30, 18.80, 17.00, 14.08, 10.40, 6.46. The swelling effect in this case is reduced compared to the Euclidean interpolation. Third row: affine-invariant interpolation. Fourth row: Log-Euclidean interpolation. Bottom: Log-Cholesky geometric interpolation. The associated determinants for the last three interpolations are the same: 5.40, 5.50, 5.60, 5.70, 5.80, 5.91, 6.01, 6.12, 6.23, 6.34, 6.46. There is no swelling effect observed for affine-invariant, Log-Euclidean, and Log-Cholesky interpolation.

5. Concluding remark. We have constructed a new Lie group structure on SPD matrices via Cholesky decomposition and a bi-invariant metric on it, termed Log-Cholesky metric. Such structure and metric have the advantages of the Log-Euclidean metric and affine-invariant metric. In addition, it has a simple and closed form for Fréchet averages and parallel transport along geodesics. For all of these metrics, Fréchet averages have the same determinant and do not have swelling effect to which both the Euclidean metric and the classic Cholesky distance are subject. Computationally, it is much faster than its two counterparts, the Log-Euclidean metric and the affine-invariant metric. For computation of parallel transport, it could be approximately 45 times faster than the Log-Euclidean metric and 4 times faster than the affine-invariant one. The Log-Cholesky metric is also numerically stabler than these two metrics.

In practice, which metric to choose may depend on the context of applications while the presented Log-Cholesky metric offers a choice alternative to existing metrics

like the Log-Euclidean metric and the affine-invariant metric. For big data sets, the advantage of the Log-Cholesky metric in computation is attractive. However, for applications like [5] to which the congruence invariance property is central, the affine-invariant metric is recommended, since numerical experiments suggest that both the Log-Euclidean and Log-Cholesky metrics do not have the congruence invariance property that is enjoyed by the affine-invariant metric.

One shall also note that the Log-Cholesky metric can be equivalently formulated in terms of upper triangular matrices. In the future, we plan to investigate other properties of Log-Cholesky means, e.g., their anisotropy and relation to other geometric means or Fréchet means. We also plan to compare the performance of various metrics in the study of brain functional connectivities which are often characterized by SPD matrices.

Acknowledgment. The author thanks Dr. Peng Ding and Dr. Qiang Sun for providing a potential application of the proposed Log-Cholesky framework when the latter approached the author for a geometric interpretation of their statistical model.

REFERENCES

- [1] K. AFTAB AND R. HARTLEY, *Lq averaging for symmetric positive-definite matrices*, in Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2013.
- [2] T. ANDO, C.-K. LI, AND R. MATHIAS, *Geometric means*, Linear Algebra Appl., 385 (2004), pp. 305–334.
- [3] M. ARNAUDON, F. BARBARESCO, AND L. YANG, *Riemannian medians and means with applications to radar signal processing*, IEEE J. Selected Topics Signal Process., 7 (2013), pp. 595–604.
- [4] V. ARSIGNY, P. FILLARD, X. PENNÉC, AND N. AYACHE, *Geometric means in a novel vector space structure on symmetric positive-definite matrices*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 328–347.
- [5] A. BARACHANT, S. BONNET, M. CONGEDO, AND C. JUTTEN, *Multiclass brain-computer interface classification by Riemannian geometry*, IEEE Trans. Biomed. Engrg., 59 (2012), pp. 920–928.
- [6] R. BHATIA, *Positive Definite Matrices*, Princeton University Press, 2007.
- [7] R. BHATIA, T. JAIN, AND Y. LIM, *On the Bures–Wasserstein distance between positive definite matrices*, Expositiones Mathematicae, 37 (2018), pp. 165–191.
- [8] R. BHATTACHARYA AND V. PATRANGENARU, *Large sample theory of intrinsic and extrinsic sample means on manifolds. I*, Ann. Stat., 31 (2003), pp. 1–29.
- [9] R. CASEIRO, J. F. HENRIQUES, P. MARTINS, AND J. BATISTA, *A nonparametric Riemannian framework on tensor field with application to foreground segmentation*, Pattern Recognition, 45 (2012), pp. 3997–4017.
- [10] Z. CHEBBI AND M. MOAKHER, *Means of Hermitian positive-definite matrices based on the log-determinant α -divergence function*, Linear Algebra Appl., 436 (2012), pp. 1872–1889.
- [11] I. L. DRYDEN, A. KOLOYDENKO, AND D. ZHOU, *Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging*, Ann. Appl. Stat., 3 (2009), pp. 1102–1123.
- [12] R. L. EUBANK AND S. WANG, *The equivalence between the Cholesky decomposition and the Kalman filter*, Amer. Stat., 56 (2002), pp. 39–43.
- [13] P. FILLARD, V. ARSIGNY, X. PENNÉC, K. M. HAYASHI, P. M. THOMPSON, AND N. AYACHE, *Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines*, NeuroImage, 34 (2007), pp. 639–650.
- [14] P. T. FLETCHER AND S. JOSHIB, *Riemannian geometry for the statistical analysis of diffusion tensor data*, Signal Processing, 87 (2007), pp. 250–262.
- [15] K. J. FRISTON, *Functional and effective connectivity: a review*, Brain Connectivity, 1 (2011), pp. 13–36.
- [16] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, 1996.
- [17] I. GRUBIŠIĆ AND R. PIETERSZ, *Efficient rank reduction of correlation matrices*, Linear Algebra Appl., 422 (2007), pp. 629–653.

- [18] J. GUILLEMINOT AND C. SOIZE, *Generalized stochastic approach for constitutive equation in linear elasticity: a random matrix model*, Int. J. Numer. Methods Engrg., 90 (2012), pp. 613–635.
- [19] S. HELGASON, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Amer. Math. Soc., 2001.
- [20] R. HOSSEINI AND S. SRA, *Matrix manifold optimization for Gaussian mixtures*, in Proceedings of the NIPS, 2015.
- [21] X. HUA, Y. CHENG, H. WANG, Y. QIN, Y. LI, AND W. ZHANG, *Matrix CFAR detectors based on symmetrized Kullback-Leibler and total Kullback-Leibler divergences*, Digital Signal Processing, 69 (2017), pp. 106–116.
- [22] S. A. HUETTEL, A. W. SONG, AND G. MCCARTHY, *Functional Magnetic Resonance Imaging*, 2nd ed., Sinauer Associates, 2008.
- [23] H. J. KIM, J. X. B. C. VEMURI, AND V. SINGH, *Manifold-valued Dirichlet processes*, in Proceedings of the 32nd International Conference on Machine Learning, vol. 37, 2015.
- [24] S. LANG, *Differential and Riemannian Manifolds*, Springer, New York, 1995.
- [25] D. LE BIHAN, *Molecular diffusion nuclear magnetic resonance imaging*, Magnetic Resonance Quarterly, 7 (1991), pp. 1–30.
- [26] J. M. LEE, *Manifolds and Differential Geometry*, Graduate Studies in Mathematics, Amer. Math. Soc., 2009.
- [27] C. LENGLLET, M. ROUSSON, R. DERICHE, AND O. FAUGERAS, *Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing*, J. Math. Imaging Vision, 25 (2006), pp. 423–444.
- [28] E. MASSART AND P.-A. ABSIL, *Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices*, tech. report, ICTEAM Institute, UCLouvain, 2018.
- [29] J. MILNOR, *Morse Theory*, Princeton University Press, 1963.
- [30] M. MOAKHER, *A differential geometry approach to the geometric mean of symmetric positive-definite matrices*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 735–747.
- [31] M. MOAKHER, *On the averaging of symmetric positive-definite tensors*, J. Elasticity, 82 (2006), pp. 273–296.
- [32] D. OSBORNE, V. PATRANGENARU, L. ELLINGSON, D. GROISSER, AND A. SCHWARTZMAN, *Non-parametric two-sample tests on homogeneous Riemannian manifolds, Cholesky decompositions and diffusion tensor image analysis*, J. Multivariate Anal., 119 (2013), pp. 163–175.
- [33] X. PENNEC, *Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements*, J. Math. Imaging Vision, 25 (2006), pp. 127–154.
- [34] X. PENNEC, P. FILLARD, AND N. AYACHE, *A Riemannian framework for tensor computing*, Int. J. Comput. Vision, 66 (2006), pp. 41–66.
- [35] Y. RATHI, A. TANNENBAUM, AND O. MICHAJLOVICH, *Segmenting images on the tensor manifold*, in Proceedings of Computer Vision and Pattern Recognition, 2007.
- [36] J.-B. SCHIRATTI, S. ALLASSONNIÈRE, O. COLLIOT, AND S. DURRLEMAN, *A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations*, J. Mach. Learn. Res., 18 (2017), pp. 1–33.
- [37] S. SRA, *Positive definite matrices and the S-divergence*, Proc. Amer. Math. Soc., 144 (2016), pp. 2787–2797.
- [38] B. VANDEREYCKEN, P.-A. ABSIL, AND S. VANDEWALLE, *A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank*, IMA J. Numer. Anal., 33 (2013), pp. 481–514.
- [39] Z. WANG, B. C. VEMURI, Y. CHEN, AND T. H. MARECI, *A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI*, IEEE Trans. Medical Imaging, 23 (2004), pp. 930–939.
- [40] Y. YUAN, H. ZHU, W. LIN, AND J. S. MARRON, *Local polynomial regression for symmetric positive definite matrices*, J. Royal Statistical Society: Series B (Statistical Methodology), 74 (2012), pp. 697–719.
- [41] M. J. ZEESTRATEN, I. HAVOUTIS, J. AO SILVÉRIO, S. CALINON, AND D. G. CALDWELL, *An approach for imitation learning on Riemannian manifolds*, IEEE Robotics Automat. Lett., 2 (2017), pp. 1240–1247.