Published in final edited form as:

Int Stat Rev. 2016 December ; 84(3): 456–486. doi:10.1111/insr.12113.

Lognormal Distributions and Geometric Averages of Symmetric Positive Definite Matrices

Armin Schwartzman

Department of Statistics, North Carolina State University

Abstract

This article gives a formal definition of a lognormal family of probability distributions on the set of symmetric positive definite (SPD) matrices, seen as a matrix-variate extension of the univariate lognormal family of distributions. Two forms of this distribution are obtained as the large sample limiting distribution via the central limit theorem of two types of geometric averages of i.i.d. SPD matrices: the log-Euclidean average and the canonical geometric average. These averages correspond to two different geometries imposed on the set of SPD matrices. The limiting distributions of these averages are used to provide large-sample confidence regions and two-sample tests for the corresponding population means. The methods are illustrated on a voxelwise analysis of diffusion tensor imaging data, permitting a comparison between the various average types from the point of view of their sampling variability.

1 Introduction

A scalar random variable is called lognormal if its logarithm is normally distributed. Analogously, a random symmetric positive definite (SPD) matrix may be called lognormal if its matrix logarithm, a random symmetric matrix, has a matrix-variate normal distribution. The resulting distribution on the set of SPD matrices can be naturally called an SPD-matrix-variate lognormal distribution.

Just like the scalar lognormal distribution may be obtained as the limiting distribution of the geometric average of i.i.d. random variables (Johnson and Kotz, 1970), it is shown here that two types of SPD-matrix-variate lognormal distributions arise as the limiting distributions of two types of geometric averages of i.i.d. random SPD matrices. The main goal of this paper is to justify and make explicit the definition of such distributions. Moreover, it is shown how these distributions can be used to construct confidence regions (CR's) for those geometric averages and perform matrix-variate two-sample tests in data analysis.

The first geometric average considered in this paper, the log-Euclidean average, is obtained by averaging the SPD matrices after a matrix-log transformation, and then transforming back by the matrix exponential (Arsigny et al., 2006, 2007; Schwartzman, 2006). Application of the multivariate central limit theorem (CLT) in the log domain leads directly to the definition of an SPD-matrix-variate log-Euclidean lognormal distribution, so that the log-Euclidean average can be regarded as approximately lognormal for large sample sizes. The second geometric average, called here canonical geometric average (Batchelor et al., 2005; Moakher, 2005; Bhatia and Holbrook, 2005; Pennec, 2006a; Lenglet et al., 2006;

Schwartzman, 2006; Fletcher and Joshi, 2007; Zhu et al., 2009; Lawson and Lim, 2013), is the intrinsic average associated with the so-called canonical or affine-invariant metric defined on the set of SPD matrices (Skovgaard, 1984; Terras, 1987; Calvo and Oller, 1991; Kobayashi and Nomizu, 1996; Lang, 1999; Smith, 2005; Moakher and Zéraï, 2011). This case requires application of a Riemannian manifold version of the CLT (Bhattacharya and Patrangenaru, 2005), and the limiting distribution leads to the definition of a second type of lognormal distribution, which we call canonical. The log-Euclidean and canonical lognormal distributions coincide when the mean parameter in both is a multiple of the identity matrix.

As opposed to former descriptions in the matrix analysis and image analysis literatures, the presentation here takes a statistical modeling perspective, reviewing and establishing properties of these averages such as uniqueness, equivariance, consistency and limiting distribution. For context, the two geometric averages are contrasted with the simpler arithmetic or Euclidean average of SPD matrices, which is also consistent but asymptotically normal. Their different properties are described in terms of the geometry of the set of SPD matrices.

In classical multivariate statistics, the most common probability model for SPD matrices has been the Wishart distribution. Related parametric models for SPD matrices include the multivariate gamma, inverse Wishart, and inverse multivariate gamma, as well as noncentral versions of these distributions (Anderson, 2003; Gupta and Nagar, 2000; Muirhead, 1982). These models assume that the observations are real-valued multivariate vectors and have as a goal to model their population or sample covariance matrix. The idea of using a matrix logarithm of SPD matrices to flexibly model covariance matrices in a similar multivariate context was explored by Leonard and Hsu (1992) and Chiu et al. (1996). In the context of harmonic analysis, CLTs for convolutions of SPD matrices were studied by Terras (1987).

Today, new data exist where the observations themselves (after pre-processing) take the form of SPD matrices. In Diffusion Tensor Imaging (DTI), a special modality of magnetic resonance imaging, each voxel contains a random 3×3 SPD matrix called diffusion tensor (Basser and Pierpaoli, 1996; LeBihan et al., 2001; Bammer et al., 2002). In the brain, the diffusion tensor describes the local spatial pattern of water diffusion, its eigenvalues related to the type of tissue and its health, and its eigenvectors related to the underlying spatial direction of neural fibers. In tensor-based morphometry (TBM), brain images are spatially registered to a template and the transformation is summarized at each voxel by a 3×3 SPD matrix called deformation tensor (Joshi et al., 2007; Leporé et al., 2008). Both the log-Euclidean and canonical geometric averages of SPD matrices, as well as the simpler Euclidean average, have been used in DTI data (Batchelor et al., 2005; Fillard et al., 2007; Schwartzman, 2006; Fletcher and Joshi, 2007; Zhu et al., 2009; Schwartzman et al., 2010) and TBM data (Leporé et al., 2008). Arguments have been made in favor and against the use of each one of these averages based on their mathematical, computational and interpretive properties (Arsigny et al., 2006; Fillard et al., 2007; Pasternak et al., 2010). Like Pasternak et al. (2010), this paper considers these averages from the point of view of their variability.

First, the CLTs in this paper enable the construction of asymptotic CR's for these averages. Visualization of these 6-dimensional regions is difficult, so in order to capture the largest

variation inside the CR, here we consider the display of the two ellipsoids at the boundary of the CR along its first principal axis. Interestingly, finding such boundary points for the canonical geometric average requires solving a special case of the continuous-time algebraic Ricatti equation, whose solution is in itself a canonical geometric average.

The CLTs in this paper also lead to asymptotic two-sample tests for geometric means of SPD matrices. The tests are set up to allow unequal covariance matrices between the groups. As a specific data example, the three types of average are evaluated in a voxelwise comparison of DTI images corresponding to two groups of 10-year-old children (12 boys and 22 girls) (Dougherty et al., 2007; Schwartzman et al., 2010). The analysis and adjoining simulations show that all three averages have very similar values and are almost equally powerful in detecting voxelwise differences between the groups.

The parametric approach followed in this paper contrasts and complements the nonparametric approach of Ellingson et al. (2012); Osborne et al. (2013), which uses bootstrapping to establish significance in two-sample tests of canonical geometric means but is restricted to analysis of a single voxel because of computational complexity. In contrast, the parametric approach followed here can be easily applied in parallel to thousands of voxels in the entire brain volume.

The rest of the paper is organized as follows. Section 2 gives the general context of the paper in terms of the Euclidean average. Section 3 introduces the log-Euclidean average and the log-Euclidean lognormal distribution. Section 4 presents the canonical geometric average and the resulting canonical lognormal distribution. Section 5 presents the DTI example, concluding with a discussion in Section 6. Relevant auxiliary material appear in Appendix A and proofs are placed in Appendix B.

2 Preliminaries

2.1 The set of SPD matrices

It is useful to think of the set $\text{Sym}^+(p)$ of $p \times p$ real SPD matrices in relation to the set $\text{Sym}(p)$ of $p \times p$ real symmetric matrices in three different ways. First, $\text{Sym}^+(p)$ is a simply connected open subset of $\text{Sym}(p)$, whose boundary is the set of $p \times p$ symmetric matrices with one or more zero eigenvalues. It is a convex cone because $X_1, X_2 \in \text{Sym}^+(p)$ implies $a_1 X_1 + a_2 X_2 \in \text{Sym}^+(p)$ for any scalars $a_1, a_2 > 0$. However, it is not a vector space; e.g. the negative of a SPD matrix and some linear combinations of SPD matrices are not SPD. Figure 1 shows the shape of the cone for $p = 2$ by embedding $\text{Sym}^+(2)$ in $\mathbb{R}^3$.

Second, $\text{Sym}^+(p)$ and $\text{Sym}(p)$ have a one-to-one correspondence via the matrix exponential and logarithm

$$\text{Sym}^+(p) = \exp[\text{Sym}(p)], \quad \text{Sym}(p) = \log[\text{Sym}^+(p)], \quad (1)$$

(see Appendix A.2). Third, $\text{Sym}^+(p)$ is a differentiable manifold whose tangent space at any point can be identified as a copy of $\text{Sym}(p)$. Furthermore, it can be given a Riemannian

manifold structure by defining a metric, which in this paper will be Euclidean, log-Euclidean, or canonical (affine-invariant) (see Appendix A.3 and A.4). These relationships will play various roles in the development below.

A useful geometrical representation of the elements of $\text{Sym}^+(p)$ themselves is the association of $X \in \text{Sym}^+(p)$ with the ellipsoid in p

$$\mathcal{E}(X) = \left\{ w \in p : w' X^{-1} w = 1 \right\}. \quad (2)$$

This ellipsoid can be thought of as an isosurface of a multivariate normal distribution in p with covariance X . This representation is useful for visualization of X as the semiaxes of $\mathcal{E}(X)$ correspond to the eigenvectors of X scaled by the square root of the eigenvalues. The representation of some elements of $\text{Sym}(2)$ as ellipses is illustrated in Figures 1 and 2.

2.2 The Euclidean average of SPD matrices and the symmetric-matrix normal distribution

Before presenting the geometric averages, it is helpful to first consider the Euclidean average for reference. Given SPD matrices $X_1, \dots, X_n \in \text{Sym}^+(p)$, the convexity of $\text{Sym}^+(p)$ implies that the Euclidean average

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (3)$$

is also an element of $\text{Sym}^+(p)$ (Figure 1). Suppose $X_1, \dots, X_n \in \text{Sym}^+(p)$ are pairwise i.i.d. with Euclidean mean $E(X_i) = M \in \text{Sym}^+(p)$ (capital μ) and Euclidean covariance $\text{Cov}(\text{vecd}(X_i)) = \Sigma \in \text{Sym}^+(q)$. Here, $\text{vecd}(\cdot)$ can be in general any vectorization operator, but we prefer the specific operator defined by (36) in Appendix A.1. Then by the multivariate CLT, as $n \rightarrow \infty$,

$$\bar{X} \rightarrow M, \quad \sqrt{n} \text{vecd}(\bar{X} - M) \Rightarrow N(0, \Sigma) \quad (4)$$

in probability and in distribution, respectively. Thus $\bar{X} \in \text{Sym}^+(p)$ is a consistent estimator of M and, for large n , is approximately normally distributed in $\text{Sym}(p)$ with mean M and covariance Σ/n (Σ can be estimated consistently by the sample covariance matrix of $\text{vecd}(X_1), \dots, \text{vecd}(X_n)$ if finite fourth order moments are also assumed).

In this first view, the SPD matrices $X_1, \dots, X_n \in \text{Sym}^+(p)$ are treated as if they were mere symmetric matrices. Because the set $\text{Sym}^+(p)$ is convex, the distribution of $\bar{X}$ is always supported on a subset of $\text{Sym}^+(p)$. However, as n increases, the distribution of $\bar{X}$ becomes more concentrated around M and the SPD constraints at the boundary of $\text{Sym}^+(p)$ become asymptotically irrelevant. Since the normal distribution is not defined on $\text{Sym}^+(p)$ for any n because of the boundary constraints, the asymptotic normal distribution of $\bar{X}$ is technically

defined on the tangent space to $\text{Sym}^+(p)$ at M , equal to $\text{Sym}(p)$. With some abuse of notation, we write this as

$$\bar{X} \sim N(N, \Sigma/n), \quad (5)$$

where we use the notational shorthand

$$X \sim N(M, \Sigma) \iff \text{vecd}(X) \sim N(\text{vecd}(M), \Sigma). \quad (6)$$

The specific choice of the vectorization operator $\text{vecd}(\cdot)$ given by (36) gives an elegant definition of the symmetric-matrix-variate normal distribution above (Schwartzman et al., 2008), which in the case of identity covariance reduces to the random matrix model called Gaussian Orthogonal Ensemble (Mehta, 1991).

Based on (4), a confidence region (CR) for M of asymptotic level α is given by the q -dimensional ellipsoid

$$CR_{n,\alpha}(\bar{X}) = \left\{ M \in \text{Sym}^+(p) : n \text{vecd}(\bar{X} - M)' \bar{\Sigma}^{-1} \text{vecd}(\bar{X} - M) \leq \chi_{q,1-\alpha}^2 \right\}, \quad (7)$$

where $\chi_{q,1-\alpha}^2$ is the $1 - \alpha$ quantile of the χ^2 distribution with q degrees of freedom and $\hat{\Sigma}$ is a consistent estimator of Σ such as the sample covariance matrix of $\text{vecd}(X_1), \dots, \text{vecd}(X_n)$.

2.3 A two-sample test of Euclidean means

In data analysis, it is often of interest to compare two sets of SPD matrices. Suppose we observe two independent random samples $X_1, \dots, X_{n_1} \in \text{Sym}^+(p)$ i.i.d. with Euclidean mean M_1 and Euclidean covariance Σ_1 , and $X_{n_1+1}, \dots, X_n \in \text{Sym}^+(p)$ i.i.d. with Euclidean mean M_2 and Euclidean covariance $\text{Cov}(\text{vecd}(X_j)) = \Sigma_2$. We assume that the covariances Σ_1 and Σ_2 are not equal because it is required for analysis of the data. Let $\bar{X}_1$ and $\bar{X}_2$ be the respective Euclidean averages, so that

$$\bar{X}_1 \sim N(M_1, \Sigma_1/n_1), \quad \bar{X}_2 \sim N(M_2, \Sigma_2/n_2),$$

for large n_1 and n_2 , as in (5). Testing the null hypothesis $H_0 : M_1 = M_2$ is a matrix-variate version of the Behrens-Fisher problem. A solution used by Schwartzman et al. (2010) is to use the scaled Hotelling's two-sample T-squared statistic

$$T_E = \frac{f - q + 1}{qf} d' S^{-1} d \underset{H_0}{\sim} F(q, f - q + 1), \quad (8)$$

where $d = \text{vecd}(\bar{X}_1 - \bar{X}_2)$, $S = S_1/n_1 + S_2/n_2$, and S_1 and S_2 are the sample Euclidean covariance matrices of the two samples in vectorized form, respectively. Here, f is the approximate number of degrees of freedom given by (Yao, 1965; Mardia, 1979)

$$\frac{1}{f} = \frac{1}{n_1 - 1} \left(\frac{d' S^{-1} S_1 S^{-1} d}{n_1 d' S^{-1} d} \right)^2 + \frac{1}{n_2 - 1} \left(\frac{d' S^{-1} S_2 S^{-1} d}{n_2 d' S^{-1} d} \right)^2. \quad (9)$$

3 The log-Euclidean average and the SPD-matrix log-Euclidean lognormal distribution

3.1 The log-Euclidean average of SPD matrices

A second view of $\text{Sym}^+(p)$ is as the image of $\text{Sym}(p)$ via the matrix exponential. Just like positive scalar data is often log-transformed before analysis, the analogous approach to analysis of SPD matrices is to transform them into symmetric matrices by a matrix log transformation (Arsigny et al., 2006, 2007; Schwartzman, 2006). Given SPD matrices $X_1, \dots, X_n \in \text{Sym}^+(p)$, mapping them to $\text{Sym}(p)$ via the matrix logarithm, computing their Euclidean average there and then mapping the result back to $\text{Sym}^+(p)$ results in the log-Euclidean average

$$\hat{X} = \exp \left(\frac{1}{n} \sum_{i=1}^n \log X_i \right). \quad (10)$$

The symbol $\hat{X}$ reflects that this is a “curved” version of the Euclidean average (Figure 1).

Given a random SPD matrix X with probability measure Q on $\text{Sym}^+(p)$, we can define the log-Euclidean mean

$$M = \exp [E(\log X)], \quad (11)$$

where the expectation is taken with respect to the measure Q , so that the log-Euclidean average (10) corresponds to the empirical measure $\hat{Q}_n = (1/n) \sum_{i=1}^n \delta(X_i)$. A direct consequence of the analytical property (40) in Appendix A.2 is that the log-Euclidean mean and the log-Euclidean average are equivariant under similarity transformations and matrix exponentiation, including inversion. Let $GL(p)$ denote the set of $p \times p$ real invertible matrices (general linear group) and let $O(p) \subset GL(p)$ denote the set of $p \times p$ orthogonal matrices. If X is a random SPD matrix with log-Euclidean mean M , then:

1. The log-Euclidean mean of AXA^{-1} is AMA^{-1} , for $A \in GL(p)$.
2. The log-Euclidean mean of $sRXR'$ is $sRMR'$, for $s > 0$ and $R \in O(p)$.
3. The log-Euclidean mean of X^a is M^a , for $a \in \mathbb{R}$.

Similarly, given SPD matrices $X_1, \dots, X_n$ with log-Euclidean average (10):

1. The log-Euclidean average of $AX_1A^{-1}, \dots, AX_nA^{-1}$ is $A\hat{X}A^{-1}$, for $A \in GL(p)$.
2. The log-Euclidean average of $sRX_1R', \dots, sRX_nR'$ is $sR\hat{X}R'$, for $s > 0$ and $R \in O(p)$.
3. The log-Euclidean average of $X_1^a, \dots, X_n^a$ is $\hat{X}^a$, for $a \in \mathbb{R}$.

3.2 The SPD-matrix log-Euclidean lognormal distribution

Suppose now that $X_1, \dots, X_n \in \text{Sym}^+(p)$ are pairwise i.i.d. SPD matrices with log-Euclidean mean $M \in \text{Sym}(p)$ (11) and log-Euclidean covariance $\text{Cov}(\text{vecd}(\log X_i)) = \Sigma \in \text{Sym}^+(q)$. Let $\hat{X}$ denote their log-Euclidean average (10). Then, applying the multivariate CLT to the transformed matrices $\log X_1, \dots, \log X_n$ gives that, as $n \rightarrow \infty$,

$$\hat{X} \rightarrow M, \quad \sqrt{n} \text{vecd}(\log \hat{X} - \log M) \Rightarrow N(0, \Sigma) \quad (12)$$

in probability and in distribution, respectively. Thus $\hat{X} \in \text{Sym}^+(p)$ is a consistent estimator of the log-Euclidean population mean (11) (Σ can be estimated consistently by the sample covariance matrix of the vectors $\text{vecd}(\log X_1), \dots, \text{vecd}(\log X_n)$ if finite fourth order moments are also assumed).

More interestingly, the log-Euclidean average $\hat{X}$ is asymptotically lognormal in the sense that, for large n , $\log \hat{X}$ is approximately normally distributed in $\text{Sym}(p)$ with mean $\log M$ and covariance Σ/n . This leads to the first definition of the lognormal distribution.

Definition 1—We say that $X \in \text{Sym}^+(p)$ has a SPD-matrix-variate log-Euclidean lognormal distribution with log-Euclidean mean $M \in \text{Sym}^+(p)$ and log-Euclidean covariance $\Sigma \in \text{Sym}^+(q)$, denoted $X \sim LN_{LE}(M, \Sigma)$, if $\log X \sim N(\log M, \Sigma)$.

Definition 1 has the interpretation that the parameter M is precisely the log-Euclidean mean (11) of the lognormal variable X . Based on (12), we can write the approximate distribution of the log-Euclidean average for large n as

$$\hat{X} \sim LN_{LE}(M, \Sigma/n). \quad (13)$$

To generate random lognormal SPD matrices $X \sim LN_{LE}(M, \Sigma)$, the recipe is to generate random normal symmetric matrices $Y \sim N(\log M, \Sigma)$ and then apply the matrix exponential $X = \exp(Y)$. An illustration for $p = 2$ is shown in Figure 2. The representation of the random matrices and their log-Euclidean mean and average as ellipses via (2) nicely visualizes their positive definiteness and variability. However, it is less useful for representing CR's, as explained below.

The following result helps establish the density of the SPD-matrix-variate log-Euclidean lognormal distribution. The proof is given in Appendix B.

Proposition 1—Let $\lambda_1 \dots \lambda_p > 0$ denote the eigenvalues of $X \in \text{Sym}^+(p)$. The Jacobian of the transformation $Y = \log X \in \text{Sym}(p)$ is given by

$$J(X) = \mathcal{J}(Y \rightarrow X) = \frac{1}{|X|} \prod_{i < j} g(\lambda_i, \lambda_j), \quad (14)$$

where

$$g(\lambda_i, \lambda_j) \begin{cases} (\log \lambda_i - \log \lambda_j) / (\lambda_i - \lambda_j), & \lambda_i > \lambda_j \\ 1/\lambda_i, & \lambda_i = \lambda_j. \end{cases} \quad (15)$$

Making the change of variable $Y = \log X$ in Definition 1 and applying Theorem 1 gives that the density of $X \sim \text{LN}_{\text{LE}}(M, \Sigma)$ is

$$f(X; M, \Sigma) = \frac{J(X)}{(2\pi)^{q/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} \text{vecd}(\log X - \log M)' \Sigma^{-1} \text{vecd}(\log X - \log M) \right) \quad (16)$$

with respect to Lebesgue measure on $\text{Sym}^+(p)$ (as defined by vecd), where $\mathcal{J}(X)$ is the Jacobian (14). An interesting special case occurs when $M = aI_p$, $a > 0$, and Σ is an orthogonally invariant covariance matrix with respect to the operator $\text{vecd}(\cdot)$ (Schwartzman et al., 2008), e.g. $\Sigma = \sigma^2 I_q$. In that case the density (16) becomes a function of the eigenvalues of X only and is orthogonally invariant in the sense that $f(X) = f(RXR')$ for any $R \in O(p)$.

As an example of what the lognormal density looks like, Figure 3 shows the construction of a log-Euclidean lognormal density in $\text{Sym}^+(2)$ with parameters

$$M = \exp \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0 \end{pmatrix} = \begin{pmatrix} 2.05 & 0.29 \\ 0.29 & 1.03 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0.25 & 0.05 & 0 \\ 0.05 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}. \quad (17)$$

Panel (a) shows two views of the density of the normal vector $y = \text{vecd}(Y) \sim N(\text{vecd}(\log M), \Sigma)$. Applying the transformation $X = \exp(Y)$ yields the density of the vector $x = \text{vecd}(X)$, shown in two views in Panel (b). The gridlines in Panel (a) have been mapped directly to Panel (b) in order to illustrate the deformative effect of the matrix exponential. Notice the shape of the $\text{Sym}^+(2)$ cone on the right column.

Finally, another nice property of the log-Euclidean lognormal distribution is that, if $X_1, \dots, X_n$ are i.i.d. $\text{LN}_{\text{LE}}(M, \Sigma)$, then the MLEs of M and Σ are given, respectively, by the log-Euclidean average and sample covariance matrix

$$\hat{M} = \hat{X} = \exp \left(\frac{1}{n} \sum_{i=1}^n \log X_i \right), \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \text{vecd} \left(\log X_i - \log \hat{X} \right) \text{vecd} \left(\log X_i - \log \hat{X} \right)' \quad (18)$$

3.3 A confidence region for the log-Euclidean mean

Based on (13), a CR for the log-Euclidean mean of asymptotic level α is given by

$$CR_{n,\alpha}(\hat{X}) = \left\{ M \in \text{Sym}^+(p) : n \text{vecd} \left(\log \hat{X} - \log M \right)' \hat{\Sigma}^{-1} \text{vecd} \left(\log \hat{X} - \log M \right) \leq \chi_{q,1-\alpha}^2 \right\}, \quad (19)$$

where $\chi_{q,1-\alpha}^2$ is the $1 - \alpha$ quantile of the χ^2 distribution with q degrees of freedom and $\hat{\Sigma}$ is a consistent estimator of Σ such as the sample covariance matrix of $\text{vecd}(\log X_1), \dots, \text{vecd}(\log X_n)$.

An example of such a CR is shown in Figure 2b. Note that the 3-dimensional ellipsoid determined by (19) in the log domain gets distorted after inverting the log into the original domain, so that the log-Euclidean average does not correspond to the Euclidean center of the CR but rather to its log-Euclidean center. The points on the CR mesh were obtained by solving for M in (19) using the principal components of $\hat{\Sigma}$ as basis. Particularly, marked in blue are the two extreme points along the first principal component of the CR (computed in the log domain), computed by inverting (19) according to

$$\exp \left[\log \hat{X} \pm \text{vecd}^{-1} \left(\lambda_1 V_1 \chi_{q,1-\alpha}^2 / n \right) \right],$$

where λ_1, V_1 are the first eigenvalue and first eigenvector of the covariance matrix $\hat{\Sigma}$. These two points are also represented as ellipses in Figure 2a. It is worth emphasizing that the shaded region in Figure 2a is not a CR because it only captures the variability along the first principal component. While the true mean is not contained in the shaded region in Figure 2a, it is indeed contained in the CR in Figure 2b. Other visualization schemes may also be devised, for instance, by choosing points on the boundary of the CR that are multiples of the log-Euclidean average or whose corresponding ellipses have the largest and smallest plotted area.

3.4 A two-sample test of log-Euclidean means

Suppose we observe two independent random samples $X_1, \dots, X_{n_1} \in \text{Sym}^+(p)$ i.i.d. with log-Euclidean mean M_1 and log-Euclidean covariance Σ_1 , and $X_{n_1+1}, \dots, X_n \in \text{Sym}^+(p)$ i.i.d. with log-Euclidean mean M_2 and log-Euclidean covariance Σ_2 . Let $\hat{X}_1$ and $\hat{X}_2$ denote their respective log-Euclidean averages, so that

$$\log \hat{X}_1 - \log M_1 \sim N(0, \Sigma_1/n_1), \quad \log \hat{X}_2 - \log M_2 \sim N(0, \Sigma_2/n_2),$$

for large n_1 , and n_2 , as in (13). To test the null hypothesis $H_0: M_1 = M_2$, we can use the log-Euclidean scaled Hotelling's two-sample T-squared statistic T_{LE} , computed with formulas equal to (8) and (9), but replacing d by $\text{vecd}(\log \hat{X}_1 - \log \hat{X}_2)$, and S_1 and S_2 by the sample covariance matrices of $\text{vecd}(\log X_1), \dots, \text{vecd}(\log X_{n_1})$ and $\text{vecd}(\log X_{n_1+1}), \dots, \text{vecd}(\log X_n)$, respectively, as in (18).

4 The canonical geometric average and the SPD-matrix canonical lognormal distribution

A second type of geometric average, called here canonical geometric average (Batchelor et al., 2005; Moakher, 2005; Bhatia and Holbrook, 2005; Pennec, 2006a; Lenglet et al., 2006; Schwartzman, 2006; Fletcher and Joshi, 2007; Zhu et al., 2009; Lawson and Lim, 2013), is obtained as the Fréchet average according to the so-called canonical or affine-invariant metric defined on $\text{Sym}^+(p)$ (Skovgaard, 1984; Terras, 1987; Calvo and Oller, 1991; Kobayashi and Nomizu, 1996; Lang, 1999; Smith, 2005; Moakher and Zéraï, 2011). This metric has also been called “generalized Frobenius” for its similarity with the standard Frobenius metric (Schwartzman, 2006; Osborne et al., 2013), but the term “canonical” (suggested by David Groisser from the U. of Florida, private communication) may be more appropriate because of the natural construction and many appealing properties of this metric, including the affine invariance indicated by the other name (see Appendix A.4). This was the natural information metric considered in the seminal work of Skovgaard (1984), and the canonical geometric average derived from it has been claimed to be a natural average for SPD matrices (Bhatia and Holbrook, 2005).

Like the log-Euclidean average, the canonical geometric average also leads asymptotically to a lognormal distribution, but of a different kind.

4.1 Fréchet and intrinsic means in $\text{Sym}^+(p)$

In general, the Fréchet mean of a probability measure Q on a complete metric space with distance ρ is the minimizer of the function $F(P) = \int \rho^2(P, X) Q(dX)$ (Fréchet, 1948; Bhattacharya and Patrangenaru, 2003, 2005; Pennec, 2006b). In particular, for the empirical measure $\hat{Q}_n = (1/n) \sum_{i=1}^n \delta(X_i)$ of a set of i.i.d. data points $X_1, \dots, X_n$, the minimizer of the function $\hat{F}_n(P) = (1/n) \sum_{i=1}^n \rho^2(P, X_i)$ is called the sample Fréchet mean or Fréchet average. If the measure Q is defined on a Riemannian manifold with a geodesic distance, then the Fréchet mean associated with this distance is called the intrinsic mean. If the manifold is embedded in Euclidean space, then the Fréchet mean associated with the restriction to the manifold of the Euclidean distance in the ambient space is called the extrinsic mean.

Seen in this context, the Euclidean average (3) and corresponding Euclidean mean $E(X) = \int X Q(dX)$ are respectively the unique Fréchet average and mean according to the Euclidean or Frobenius distance

$$d_E^2(P, X) = \|X - P\|^2 = \text{tr} \left[(X - P)^2 \right]. \quad (20)$$

Because of the inclusion $\text{Sym}^+(p) \subset \text{Sym}(p)$, the embedding of $\text{Sym}^+(p)$ in Euclidean space and its endowment with the Euclidean Riemannian metric are immediate, so the Euclidean average and mean are both trivially extrinsic and intrinsic.

The situation is similar for the log-Euclidean average (10) and log-Euclidean mean (11). They are respectively the unique Fréchet average and mean according to the log-Euclidean distance

$$d_{LE}^2(P, X) = \|\log X - \log P\|^2 = \text{tr} \left[(\log X - \log P)^2 \right]. \quad (21)$$

As shown by Arsigny et al. (2007), $\text{Sym}^+(p)$ can be given a complete Riemannian structure as an Abelian Lie group with group operation $(A, B) \rightarrow A \odot B = \exp(\log A + \log B)$, so that the bi-invariant log-Euclidean metric (21) is geodesic. The log-Euclidean average and log-Euclidean mean are intrinsic with respect to this metric. They are also extrinsic, because the embedding provided by the mapping (1) is flat.

4.2 The canonical geometric mean in $\text{Sym}^+(p)$

In the case of $\text{Sym}^+(p)$, the canonical geometric mean of a measure Q is defined as the intrinsic mean of Q according to the canonical geodesic distance

$$d_C^2(P, X) = \|\log(P^{-1/2} X P^{-1/2})\|^2 = \text{tr} \left\{ \left[\log(P^{-1/2} X P^{-1/2}) \right]^2 \right\}. \quad (22)$$

With this metric, $\text{Sym}^+(p)$ becomes a geodesically complete Riemannian manifold (see Appendix A.4). The canonical geometric mean is characterized by the following result.

Proposition 2—*Let Q be a probability measure on $\text{Sym}^+(p)$ such that*

$F_C(P) = \int d_C^2(P, X) Q(dX)$ is finite for $P \in \text{Sym}^+(p)$, where d_C is the canonical geodesic distance (22). Then the canonical geometric mean M is the unique solution to the equation

$$\int \log(M^{-1/2} X M^{-1/2}) Q(dX) = 0. \quad (23)$$

This result can be seen as a consequence of Theorem 2.1 in Bhattacharya and Patrangenaru (2003) or Corollary 2 in Pennec (2006b) for complete simply connected Riemannian manifolds, although the concept already appears in Oller and Corcuera (2003). The

uniqueness is guaranteed by the non-positive curvature of $\text{Sym}^+(p)$ with the canonical metric (see Section A.4 above). Equation (23) indicates the desired property that the residuals about the intrinsic mean, defined here as the image of the data points onto the tangent space at M via the Riemannian logarithmic map (given by (47) in Appendix A.4) average to zero.

In relation to (23), letting $Y = \log(M^{-1/2} X M^{-1/2})$, we may define the canonical geometric covariance matrix of the measure Q as

$$\Sigma = \text{Cov}[\text{vecd}(Y)] = \int \text{vecd}(Y) \text{vecd}(Y)' \tilde{Q}(dY), \quad (24)$$

where $\tilde{Q}$ is the measure induced by the transformation $X \rightarrow Y$ so that $\int Y \tilde{Q}(dY) = 0$.

Note that the canonical geometric mean and the log-Euclidean mean are not the same because, in general, $\log(M^{-1/2} X M^{-1/2}) \neq \log X - \log M$. The only case where equality holds for all $X \in \text{Sym}^+(p)$ is when M is a multiple of the identity matrix, i.e. $M = aI$ for some $a > 0$. Another situation when equality holds is the degenerate case where the measure Q is supported on values of X that have all the same eigenvectors, so that X and M commute (Moakher, 2005; Arsigny et al., 2007).

The canonical geometric mean has the following equivariance properties. If X is a random SPD matrix according to a probability measure Q with canonical geometric mean M , then:

1. The canonical geometric mean of AXA^{-1} is AMA^{-1} , for $A \in GL(p)$.
2. The canonical geometric mean of GXG' is GMG' , for $G \in GL(p)$.
3. The canonical geometric mean of X^{-1} is M^{-1} .

These properties parallel the invariance properties of the canonical distance (see (48) in Appendix A.4). In comparison with the log-Euclidean mean (Section 3.1), the canonical geometric mean is equivariant with respect to the group action of $GL(p)$ (see (44) in Appendix A.4), but its invariance with respect to inversion does not extend in general to other matrix powers.

4.3 The canonical geometric average in $\text{Sym}^+(p)$

For a set of i.i.d. data points $X_1, \dots, X_n \in \text{Sym}^+(p)$, their canonical geometric average, denoted here $\overset{\circ}{X}_1$, is defined as their sample intrinsic mean according to the geodesic distance (22). Applying Proposition 2 to the empirical measure $\hat{Q}_n = (1/n) \sum_{i=1}^n \delta(X_i)$ gives that the canonical geometric average is unique and satisfies

$$\frac{1}{n} \sum_{i=1}^n \log \left(\overset{\circ}{X}^{-1/2} X_i \overset{\circ}{X}^{-1/2} \right) = 0. \quad (25)$$

This condition can also be obtained by direct differentiation of the target variance function

$$\sum_{i=1}^n d_C^2(P, X_i) \quad (\text{Moakher, 2005}).$$

Unlike the Euclidean and log-Euclidean averages, the canonical geometric average in general has no closed analytical form, except for a few cases detailed in Moakher (2005). The following iterative algorithm is commonly used for numerically computing the canonical geometric average (Groisser, 2004; Pennec, 2006a; Fletcher and Joshi, 2007).

Algorithm 1

1. For $k = 0$, set an initial value $\overset{\circ}{X}^{(0)}$ such as the Euclidean average $\bar{X}$ or the log-Euclidean average $\hat{X}$. Fix $\varepsilon > 0$.

2. Evaluate

$$\bar{Y}^{-(k)} = \frac{1}{n} \sum_{i=1}^n \log \left[\left(\overset{\circ}{X}^{(k)} \right)^{-1/2} X_i \left(\overset{\circ}{X}^{(k)} \right)^{-1/2} \right], \quad \overset{\circ}{X}^{(k+1)} = \left(\overset{\circ}{X}^{(k)} \right)^{1/2} \exp \left(\bar{Y}^{-(k)} \right) \left(\overset{\circ}{X}^{(k)} \right)^{1/2}.$$

3. If $\|\bar{Y}^{-(k)}\| < \varepsilon$, stop. Otherwise, increase k by 1 and go back to Step 2.

In Algorithm 1, each iteration in Step 2 maps the data onto the tangent space at $\overset{\circ}{X}^{(k)}$ via the Riemannian log map, finds the Euclidean average $\bar{Y}^{(k)}$ there, and then maps it back onto the manifold via the Riemannian exponential map to obtain $\overset{\circ}{X}^{(k+1)}$. The stopping condition in Step 3 reflects the property (25) that the average of the residuals should converge to zero.

Because the canonical geometric average is a special case of the canonical geometric mean with the empirical measure $\hat{Q}_n$, it has the same equivariance properties. Namely, if $X_1, \dots, X_n$ are SPD matrices with canonical geometric average $\overset{\circ}{X}$, then:

1. The canonical geometric average of $AX_1A^{-1}, \dots, AX_nA^{-1}$ is $\overset{\circ}{A} \overset{\circ}{X} \overset{\circ}{A}^{-1}$, for $A \in GL(p)$.
2. The canonical geometric average of $GX_1G', \dots, GX_nG'$ is $\overset{\circ}{G} \overset{\circ}{X} \overset{\circ}{G}'$, for $G \in GL(p)$.
3. The canonical geometric average of $X_1^{-1}, \dots, X_n^{-1}$ is $\overset{\circ}{X}^{-1}$.

The consistency of the sample canonical geometric mean is given by the following result.

Proposition 3—Let Q be a probability measure on $\text{Sym}^+(p)$ such that

$F_C(P) = \int d_C^2(P, X) Q(dX)$ is finite for $P \in \text{Sym}^+(p)$, where d_C is the canonical geodesic distance (22). Then the canonical geometric average $\overset{\circ}{X}$ is a strongly consistent estimator of the canonical geometric mean M .

This result is a direct consequence of Theorem 2.3 of Bhattacharya and Patrangenaru (2003) for complete metric spaces and is understood in the sense that, given any $\varepsilon > 0$, the sample intrinsic mean will be within a distance less than ε from the true intrinsic mean for all sufficiently large sample sizes with probability 1.

4.4 The SPD-matrix canonical lognormal distribution

Like in the case of the log-Euclidean average, the asymptotic distribution of the canonical geometric average also leads to a matrix lognormal distribution on $\text{Sym}^+(p)$. To obtain this asymptotic distribution, we rely on the CLT for Riemannian manifolds given by Theorem 2.2 of Bhattacharya and Patrangenaru (2005), further discussed in Bhattacharya and Bhattacharya (2008).

Theorem 1—Let $X_1, \dots, X_n \in \text{Sym}^+(p)$ be i.i.d. SPD matrices according to a probability measure Q that is absolutely continuous with respect to the Euclidean volume measure in $\text{Sym}(p)$. Let M and $\overset{\circ}{X}$ be the canonical geometric mean and canonical geometric average, respectively. Further, define the i.i.d. random symmetric matrices

$$Y_i = \log \left(M^{-1/2} X_i M^{-1/2} \right), \quad i=1, \dots, n \quad (26)$$

with common canonical geometric covariance matrix $\Sigma = \text{Cov}(\text{vecd}(Y_1))$ (24) and assume that their moments of all orders are finite. Then,

$$\sqrt{n} \text{vecd} \left[\log \left(M^{-1/2} \overset{\circ}{X} M^{-1/2} \right) \right] \Rightarrow N \left(0, K^{-1} \Sigma K^{-1} \right) \quad (27)$$

in distribution, with

$$K = I + D' \left[\frac{1}{12} E(Y_1 \ominus Y_1)^2 + \frac{1}{720} E(Y_1 \ominus Y_1)^4 + \dots \right] D, \quad (28)$$

where D is the $p^2 \times q$ duplication matrix defined by (38) in Appendix A.1, and the symbol “ $\ominus$ ” denotes the operation $Y \ominus Y = Y \otimes I - I \otimes Y$.

The operator “ $\ominus$ ” may be called “Kronecker difference” in resemblance with the Kronecker sum operator $Y \oplus Y = Y \otimes I + I \otimes Y$. The matrix K reflects the effect of curvature on the asymptotic covariance. The proof of Theorem 1 is given in Appendix B. Note that, as opposed to the CLT for the log-Euclidean mean, finite moments are required for all orders, as these moments enter the calculation of the curvature matrix K .

As with the Euclidean and log-Euclidean averages, the limiting distribution of the canonical geometric average is defined on the tangent space at the true value of the canonical geometric mean. Theorem 1 states that the canonical geometric average $\overset{\circ}{X}$ is asymptotically

lognormal in the sense that, for large n , $\log \left(M^{-1/2} \overset{\circ}{X} M^{-1/2} \right)$ is approximately normally distributed in $\text{Sym}(p)$ with mean zero and covariance $K^{-1} \Sigma K^{-1}/n$. This leads to the second definition of the lognormal distribution.

Definition 2—We say that $X \in \text{Sym}^+(p)$ has a SPD-matrix-variate canonical lognormal distribution with canonical geometric mean $M \in \text{Sym}^+(p)$ and canonical geometric covariance $\Sigma \in \text{Sym}^+(p)$, denoted $X \sim \text{LN}_C(M, \Sigma)$, if $\log(M^{-1/2}XM^{-1/2}) \sim N(0, \Sigma)$.

Definition 2 has the interpretation that the parameter M is precisely the canonical geometric mean of the lognormal variable X because it satisfies (23). Based on (27), Definition 2 allows us to write the approximate distribution of the log-Euclidean average for large n as

$$\bar{X} \sim \text{LN}_C\left(M, K^{-1}\Sigma K^{-1}/n\right). \quad (29)$$

Note that the canonical lognormal distribution given by Definition 2 is not the same as the log-Euclidean lognormal distribution given by Definition 1 for the same reason that the log-Euclidean mean and the canonical mean are not the same. In the special case that M is a multiple of the identity matrix, i.e. $M = aI$ for some $a > 0$, then the log-Euclidean and canonical lognormal distributions coincide.

From Definition 2, the density of $X \sim \text{LN}_C(M, \Sigma)$ is obtained by the change of variable $Y = \log(M^{-1/2}XM^{-1/2})$. Applying Theorem 1 to the random variable $\tilde{X} = M^{-1/2}XM^{-1/2}$, in addition to the Jacobian $J(\tilde{X} \rightarrow X) = 1/|M|^{(p+1)/2}$ (Muirhead, 1982), gives that X has density

$$f(X; M, \Sigma) = \frac{J(\tilde{X})}{(2\pi)^{q/2} |\Sigma|^{1/2} |M|^{(p+1)/2}} \exp\left(-\frac{1}{2} \text{vecd}(\log \tilde{X})' \Sigma^{-1} \text{vecd}(\log \tilde{X})\right) \quad (30)$$

with respect to Lebesgue measure on $\text{Sym}^+(p)$ (as defined by vecd), where $J(\tilde{X})$ is given by expression (14) but replacing instead the eigenvalues $\tilde{\lambda}_1 \geq \dots \geq \tilde{\lambda}_p > 0$ of $\tilde{X} = M^{-1/2}XM^{-1/2}$, also known as the generalized eigenvalues of the pair (M, X) .

For illustration, Figure 4 shows the construction of a canonical lognormal density in $\text{Sym}^+(2)$ with the same parameters as (17). Panel (a) shows two views of the density of the normal vector $y = \text{vecd}(Y) \sim N(0, \Sigma)$. Applying the transformation $X = \exp(M^{1/2}YM^{1/2})$ yields the density of the vector $x = \text{vecd}(X)$, shown in two views in Panel (b). The gridlines in Panel (a) have been mapped directly to Panel (b) in order to illustrate the deformative effect of the exponential transformation, which in contrast to Figure 3, is affected by M . Notice the shape of the $\text{Sym}^+(2)$ cone on the right column and the additional deformation in comparison with Figure 3.

Unlike the log-Euclidean lognormal distribution, whose MLE is the log-Euclidean average, the MLE of the canonical lognormal distribution is not the canonical geometric average. To see this, consider a random sample of size n and known $\Sigma = \sigma^2 I_q$ for simplicity. The log-likelihood function, up to constants, is

$$\ell(M; X_1, \dots, X_n) = \sum_{i=1}^n \log \frac{J(M^{-1/2} X_i M^{-1/2})}{|M|^{p+1}} - \frac{1}{2\sigma^2} \sum_{i=1}^n \text{tr} \left[\log(M^{-1/2} X_i M^{-1/2}) \right]^2.$$

By definition, the canonical geometric average minimizes the last term in the likelihood. However, this is not enough to maximize the full likelihood function due to the presence of other terms that depend on M . Maximization of the likelihood, whether by analytical or numerical methods, is difficult due to the complicated form of the Jacobian.

4.5 A confidence region for the canonical geometric mean

Based on (27), a confidence region for the canonical geometric mean of asymptotic level α is given by

$$CR_{n,\alpha}(\overset{\circ}{X}) = \left\{ M \in \text{Sym}^+(p) : \text{vecd}(\overset{\circ}{Y})' \hat{K} \hat{\Sigma}^{-1} \hat{K} \text{vecd}(\overset{\circ}{Y}) \leq \chi_{q,1-\alpha}^2 \right\} \quad (31)$$

where $\overset{\circ}{Y} = \log(M^{-1/2} \overset{\circ}{X} M^{-1/2})$. The matrices $\hat{\Sigma}$ and $\hat{K}$ in (31) are consistent estimators of Σ and K , respectively. They may be obtained as

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \text{vecd}(\hat{Y}_i) \text{vecd}(\hat{Y}_i)', \quad \hat{K} = I + D' \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{12} (\hat{Y}_i \ominus \hat{Y}_i)^2 + \frac{1}{720} (\hat{Y}_i \ominus \hat{Y}_i)^4 + \dots \right] D, \quad (32)$$

where $\hat{Y}_i = \log \left(\overset{\circ}{X}^{-1/2} X_i \overset{\circ}{X}^{-1/2} \right)$, $i = 1, \dots, n$, in accordance to (26), where $\overset{\circ}{X}$ estimates M . Note that there is no need to subtract the average of the $\text{vecd}(\hat{Y}_i)$'s in the computation of the sample covariance because it is zero by (25).

Note that in (31), $\overset{\circ}{Y}$ depends on M in a nonlinear way. In the data analysis Section 5 below we will compute the extreme points along the first principal component of the CR (in the log domain). These are defined as the points M satisfying the equation

$$M^{-1/2} \overset{\circ}{X} M^{-1/2} = B_{\pm} \triangleq \exp \left[\pm \text{vecd}^{-1} \left(\lambda_1 V_1 \chi_{q,1-\alpha}^2 / n \right) \right], \quad (33)$$

where λ_1 , V_1 are the first eigenvalue and first eigenvector of the matrix $\hat{K}^{-1} \hat{\Sigma} \hat{K}^{-1}$. Equation (33) is a special case of the continuous-time algebraic Riccati equation. In general, the unique SPD solution to the equation $XA^{-1}X = B$ for $A, B \in \text{Sym}^+(p)$ is $A \# B = A(A^{-1}B)^{1/2}$ (Shurbet, 1974; Lawson and Lim, 2001; Bhatia, 2007). Interestingly, this is the

same as the canonical geometric average of A and B defined via (25) for $n = 2$ (Moakher,

2005). Therefore the solution to (33) is $M = \left[\begin{pmatrix} \overset{\circ}{X} & B_{\pm} \end{pmatrix}^{-1/2} \overset{\circ}{X} \right]^2$.

4.6 A two-sample test of canonical geometric means

Suppose we observe two independent random samples $X_1, \dots, X_{n_1} \in \text{Sym}^+(p)$ i.i.d. with canonical geometric mean M_1 and canonical geometric covariance Σ_1 , and $X_{n_1+1}, \dots, X_n \in \text{Sym}^+(p)$ i.i.d. with canonical geometric mean M_2 and canonical geometric covariance Σ_2 .

Let $\overset{\circ}{X}_1$ and $\overset{\circ}{X}_2$ denote their respective canonical geometric averages, so that

$$\begin{aligned} \log \left(M_1^{-1/2} \overset{\circ}{X}_1 M_1^{-1/2} \right) &\sim N \left(0, K_1^{-1} \Sigma_1 K_1^{-1} / n_1 \right), \\ \log \left(M_2^{-1/2} \overset{\circ}{X}_2 M_2^{-1/2} \right) &\sim N \left(0, K_2^{-1} \Sigma_2 K_2^{-1} / n_2 \right), \end{aligned}$$

for large n_1 and n_2 as in (29), where K_1 and K_2 are the curvature matrices given by expressions corresponding to (28). To test the null hypothesis $H_0 : M_1 = M_2$, we can use the canonical scaled Hotelling's two-sample T-squared statistic T_C , computed with formulas equal to (8) and (9), but replacing d by

$$\text{vecd} \left[\log \begin{pmatrix} \overset{\circ}{X}^{-1/2} & \overset{\circ}{X}_1 & \overset{\circ}{X}^{-1/2} \end{pmatrix} - \log \begin{pmatrix} \overset{\circ}{X}^{-1/2} & \overset{\circ}{X}_2 & \overset{\circ}{X}^{-1/2} \end{pmatrix} \right].$$

Here, $\overset{\circ}{X}$ is the canonical geometric average of the pooled sample of size $n = n_1 + n_2$ and serves as an estimate of the common canonical geometric mean $M = M_1 = M_2$ under the null hypothesis. Further, we replace S_1 and S_2 by $S_1 = \hat{K}_1^{-1} \hat{\Sigma}_1 \hat{K}_1^{-1}$ and $S_2 = \hat{K}_2^{-1} \hat{\Sigma}_2 \hat{K}_2^{-1}$, where $\hat{\Sigma}_1$, $\hat{K}_1$, $\hat{\Sigma}_2$ and $\hat{K}_2$ are estimated for each sample separately as in (32).

The asymptotic approximation obtained by the formulation above does not take into account the additional variability introduced by the estimation of the pooled canonical geometric mean M and the curvature matrices K_1 and K_2 . However, simulations show that this effect is negligible. Resembling the DTI data to be presented later, two independent samples of i.i.d. random SPD matrices of sizes $n_1 = 12$ and $n_2 = 22$ were generated according to the canonical lognormal distribution under two different scenarios. The null hypothesis scenario consisted of parameters

$$M_1 = M_2 = \begin{pmatrix} 0.89 & -0.02 & 0 \\ -0.02 & 0.87 & 0 \\ 0 & 0 & 0.88 \end{pmatrix}, \quad \begin{aligned} \Sigma_1 &= 0.01 \left(1_3 1_3' + I_6 \right), \\ \Sigma_2 &= 0.01 \left[10 \left(1_3 1_3' \right) + I_6 \right], \end{aligned} \quad (34)$$

where $1_3 = (1 \ 1 \ 1 \ 0 \ 0 \ 0)'$ and I_6 is the 6×6 identity matrix. Note that the mean matrices are the same in both groups but the covariance matrices are not. The non-null hypothesis scenario consisted of parameters

$$M_1 = \begin{pmatrix} 0.86 & -0.07 & -0.02 \\ -0.07 & 0.73 & 0.01 \\ -0.02 & 0.01 & 0.82 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0.93 & 0.02 & 0.01 \\ 0.02 & 0.95 & -0.01 \\ 0.01 & -0.01 & 0.92 \end{pmatrix}, \quad (35)$$

and the same covariance matrices as in (34). The parameters above correspond roughly to the sample canonical geometric means and covariances at the voxel with coordinates $(i, j) = (42, 13)$ in Figure 9 below, which is the voxel exhibiting the strongest difference between the two groups (the estimated mean in (34) is based on the pooled sample). Two-sample p-values were computed for testing equality of the Euclidean means using Euclidean averages (Section 2.3), equality of the log-Euclidean means using log-Euclidean averages (Section 3.4), and equality of the canonical geometric means using canonical geometric averages.

Figure 5 shows the p-value distribution under both scenarios for 3000 repetitions. As seen in the left panel, all three p-value distributions are very close to uniform (areas under the ecdf curves are 0.497, 0.495, 0.495, respectively). This indicates that the asymptotic approximations based on (9) are sufficiently accurate, despite the model misspecification in the Euclidean and log-Euclidean cases (the data is canonical lognormal) and the plug-in estimation of the pooled canonical geometric average in the canonical case.

The fact that the data is canonical lognormal has a stronger effect on detection power. The right panel of Figure 5 shows that the Euclidean test is slightly less powerful than the log-Euclidean and canonical tests, while the latter two are almost indistinguishable (areas under the ecdf curves are 0.996, 0.998 and 0.998, respectively). For these data, the variability around the mean is small enough that for most data points X , the two expressions $\log(M^{-1/2}XM^{-1/2})$ and $\log X - \log M$ are numerically very close.

5 A DTI data example

5.1 Data description

In this paper, we analyze a dataset consisting of 34 DTI images corresponding to 34 10-year-old children (12 boys and 22 girls). This dataset was also analyzed in Schwartzman et al. (2010) and is part of a larger observational study of brain anatomy in children (Dougherty et al., 2007). The 34 DTI images were spatially normalized to a common custom template by a smooth nonlinear deformation (Dougherty et al., 2005), obtained from corresponding T1-weighted images using six landmarks for gross registration, followed by nonlinear local deformation using SPM2 (Ashburner et al., 1999). This process resulted in each image being a voxel array of size $81 \times 106 \times 76$ in a rectangular grid with $2 \times 2 \times 2$ mm regular spacings. The coordinate axes are defined so that the x , y and z axis point respectively to the right, front and top of the brain (Figure 6). For each image, the deformation mapping was applied to the diffusion tensor (DT) map taking care to rotate the tensors to preserve their orientation (Alexander et al., 2000). Like in Schwartzman et al. (2010), analysis was restricted to voxels strictly inside the brain.

5.2 Voxelwise averages

After spatial normalization, the collection of 34 DT's across subjects at each voxel can be modeled as a collection of independent SPD matrices whose population mean and covariance depend on the voxel location. Estimation of the population mean DT image is achieved by computing SPD averages voxelwise.

Figure 6 shows the three types of voxelwise averages discussed in this paper, pooled across subjects, at a transverse slice 36 mm above the anterior commissure, an anatomical landmark commonly used for spatial normalization (Talairach convention). This is the same slice that is shown in Schwartzman et al. (2010). Because DTs are 6-dimensional objects, Figure 6 shows a meaningful reduction to 3 dimensions. One dimension, encoded in the image intensity, is captured by fractional anisotropy (FA), a function of the DT's eigenvalues. FA values near 0 represent nearly isotropic diffusion, found in the brain's water ventricles, while values near 1 represent highly anisotropic diffusion, found in tightly packed white matter neural fibers. The other two dimensions, encoded in color, are captured by the principal diffusion direction (PDD), the first eigenvector of the DT. In highly anisotropic voxels it indicates the direction of the underlying neural fibers.

At first, all three averages appear to be very similar. However, Figure 7 shows that the Euclidean and log-Euclidean averages disagree in some regions, particularly in the corpus callosum (red X-like structure in Figure 6). The differences between the log-Euclidean and canonical geometric averages are about an order of magnitude smaller.

To better visualize the differences between the three types of averages, Figure 8 shows their ellipsoidal rendering via (2) for the voxel with the largest FA difference in Figure 7. Superimposed are similar ellipsoidal renderings of the extreme points along the first principal component (in the log domain) of the corresponding 95% CR, as was done for the 2D example in Figure 2. The observed averages do exhibit visible differences in FA and orientation, but so do the extreme points of the CRs. Note that these 6-dimensional extreme points do not show the full extent of the 21-dimensional CRs but only their main mode of variation.

5.3 Two-sample tests

In DTI, it is often of interest to find voxelwise differences between two groups of DTI images, in our case boys and girls. This translates to testing at each voxel whether two groups of SPD matrices, corresponding to the diffusion tensors belonging to the two groups of subjects at that voxel, have the same mean.

A preliminary analysis of the matrix residuals with respect to the matrix averages reveals immediately that the covariance matrices in the two groups are not the same in most voxels. Similar to the analysis in Section 4.6, Figure 9 shows the p-value images for the three two-sample tests considered in this paper based on the three average types, not assuming equal covariance between the groups. The three maps are visually indistinguishable. The empirical p-value distributions of the $m = 3050$ voxels in the brain mask, shown in Figure 10, are very close to each other. The areas under the p-value ecdf are 0.5941, 0.5977 and 0.5982 respectively, giving a negligible advantage to the log-Euclidean and canonical tests. All three

tests yield the same $R = 16$ voxels with p-values below the threshold $t = 10^{-3}$, which corresponds to a false discovery rate estimate of $mt/R = 0.191$.

6 Discussion

6.1 Metrics and averages in $\text{Sym}^+(p)$

In this paper we have considered the three most common types of averages and their corresponding population means defined on $\text{Sym}^+(p)$: arithmetic or Euclidean, log-Euclidean and canonical geometric. We have seen that CLTs applied to the latter two lead to the definition of lognormal distributions on $\text{Sym}^+(p)$, analogous to the lognormal distribution for positive scalars.

These three types of averages correspond to three different metrics defined on $\text{Sym}^+(p)$. Often for data on manifolds, such as data on the sphere, Riemannian metrics are appealing because they lead to intrinsic means instead of extrinsic means. However, in the case of $\text{Sym}^+(p)$ with the Euclidean metric, the simple Euclidean mean is both intrinsic and extrinsic and has a well defined asymptotic normal distribution despite $\text{Sym}^+(p)$ not being a vector space. From this point of view, going beyond Euclidean metrics for $\text{Sym}^+(p)$ may seem unnecessary.

The DTI literature has seen a debate about which one the three metrics (Euclidean, log-Euclidean or canonical) to use in data analysis, in addition to other metrics (Dryden et al., 2009). While the canonical metric is agreed upon as the most natural metric for SPD matrices as mathematical objects (Skovgaard, 1984; Terras, 1987; Calvo and Oller, 1991; Kobayashi and Nomizu, 1996; Lang, 1999), arguments have been made in favor of the log-Euclidean metric (Arsigny et al., 2006; Whitcher et al., 2007), claiming that log-Euclidean averages are much faster to compute than canonical averages and often indistinguishable from canonical averages in real data analyses. Arguments have also been put forward against the Euclidean average for exhibiting a “swelling” effect, where its determinant may be larger than that of the SPD matrices being averaged (Arsigny et al., 2006; Fillard et al., 2007). However, (Pasternak et al., 2010) have claimed that the Euclidean average is most appropriate for DTs because it leads to a normal distribution, which is consistent with the calculation of diffusion coefficients as logarithms of the diffusion weighted imaging signal. Moreover, they claim that the scale-invariant property of the log-Euclidean and canonical geometric averages are not appropriate for DTs, leading to biases in estimation.

In this paper, we have considered the random variability of each average, interpreted here as variability between subjects rather than variability between repeated experiments, as in Pasternak et al. (2010). According to the data analysis of Section 5, agreeing with Arsigny et al. (2006), the log-Euclidean and canonical geometric averages are virtually indistinguishable, while the Euclidean and log-Euclidean averages are mostly within the expected random variation for the given sample size. Moreover, all three tests performed equally in detecting differences between the two groups of subjects in the particular data set analyzed here.

Computational considerations are also important (Groisser, 2004; Bhattacharya et al., 2012). When comparing their computational complexity, our Matlab implementation computed the Euclidean, log-Euclidean and canonical geometric averages for all 3050 voxels in the slice shown in 0.13 s, 3.95 s and 21.7 s, respectively. Times for performing two-sample tests would be proportionally longer. While these times may vary, these results make the log-Euclidean average convincingly preferable to the canonical geometric average. The choice between the Euclidean and log-Euclidean averages may be based on their desired equivariance properties (Pasternak et al., 2010), and is only of practical importance for large sample sizes.

The statistical relevance of other types of averages suggested by Dryden et al. (2009) could be analyzed using large sample techniques like the ones used in this paper.

6.2 Parametric models for random SPD matrices

The two SPD-matrix-variate lognormal families of distributions defined in this paper were obtained by applying the CLT to the log-Euclidean and canonical geometric averages of SPD matrices. We derived their explicit densities and provided CRs for the corresponding geometric means.

More generally, the SPD-matrix-variate lognormal families could be used as parametric models for SPD matrix data, either for frequentist data analysis or as flexible priors for Bayesian analysis. The lognormal distributions provide a more flexible parametric model than the Wishart / matrix Gamma or inverse Wishart / inverse matrix Gamma distributions because it has $q = p(p + 1)/2$ parameters for the mean and $q(q + 1)/2$ separate additional parameters for the covariance, while the other only have $q + 1$ parameters. As in the scalar case, the matrix lognormal distribution also has the heaviest tails.

In contrast to the CLT method used in this paper for defining a lognormal distribution, Lenglet et al. (2006) defined a normal distribution on $\text{Sym}^+(p)$ based on the maximum entropy principle developed by Pennec (2006b) for general manifolds. While algebraically similar to the canonical lognormal density (30), the density in Lenglet et al. (2006) is defined with respect to the measure induced by the canonical metric and is not lognormal. It is an interesting question whether the two densities coincide asymptotically as the variance gets smaller, as in the approximation presented in Lenglet et al. (2006). Answering this question requires a careful analysis and is left for future work.

Acknowledgments

Special thanks to Emmanuel Ben-David (Columbia U.), Victor Patrangenaru (Florida State U.) and David Groisser (U. of Florida, Gainesville) for helpful discussions, as well as Bob Dougherty (Stanford U.) for providing the DTI data. This work was partially funded by NIH grant R21EB012177.

Appendix

A Auxiliary definitions and results

A.1 The $\text{vecd}(\cdot)$ operator

For a matrix $Y \in \text{Sym}(p)$, we define the mapping of $\text{Sym}(p)$ to q with $q = p(p+1)/2$ given by

$$\text{vecd}(Y) = \left(\text{diag}(Y)', \sqrt{2} \text{ offdiag}(Y)' \right)', \quad (36)$$

where $\text{diag}(Y)$ is a $p \times 1$ column vector containing the diagonal entries of Y and $\text{offdiag}(Y)$ is a $(q-p) \times 1$ column vector containing the off-diagonal entries of Y copied from below the diagonal columnwise (or above the diagonal rowwise) (?). In the case $p=3$, $q=6$, the corresponding form

$$\text{vecd}(Y) = \left(Y_{11}, Y_{22}, Y_{33}, \sqrt{2}Y_{12}, \sqrt{2}Y_{13}, \sqrt{2}Y_{23} \right)'$$

has been used by Arsigny et al. (2006); Whitcher et al. (2007); Pasternak et al. (2010), and can be traced back to at least Mehrabadi (1995). In contrast to the usual columnwise vectorization operator $\text{vec}(\cdot)$ (Chiu et al., 1996; Gupta and Nagar, 2000) and other vectorization operators for symmetric matrices (Henderson and Searle, 1979; Magnus and Neudecker, 1980; Moakher and Zérai, 2011), the inclusion of the factor $\sqrt{2}$ in $\text{vecd}(\cdot)$ (also used by Pennec (2006a), but with the $\text{vec}(\cdot)$ ordering) yields the convenient property that the Frobenius norm of Y is the same as the Euclidean norm of $\text{vecd}(Y)$, i.e.

$$\text{tr}(Y^2) = \text{vecd}(Y)' \text{vecd}(Y). \quad (37)$$

The operator $\text{vecd}(\cdot)$ thus provides an isometric embedding of both $\text{Sym}(p)$ and $\text{Sym}^+(p)$ in q . This is the same embedding illustrated for $p=2$ in Figure 1. Also, listing the diagonals first in (36) is convenient for vectorizing diagonal matrices.

For any p , a duplication matrix D of size $p^2 \times q$ exists such that the operator $\text{vecd}(\cdot)$ is related to the usual columnwise vectorization operator $\text{vec}(\cdot)$ by

$$\text{vec}(Y) = D \text{vecd}(Y), \quad \text{vecd}(Y) = D' \text{vec}(Y). \quad (38)$$

For example, in the case $p=3$, $q=6$,

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 0 & 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

The duplication matrix D has rank q and it satisfies the following properties. For any $Y, A \in \text{Sym}(p)$,

$$\begin{aligned} D'D &= I \\ DD' \text{vec}(Y) &= \text{vec}(Y) \\ \text{vec}(AY A') &= D'(A \otimes A) D \text{vec}(Y), \end{aligned}$$

where the third property is based on the general property that

$$\text{vec}(AY B') = (B \otimes A) \text{vec}(Y) \quad (\text{Schott, 2005}).$$

A.2 Matrix exponential and logarithm

The exponential of a square matrix Y is defined as $X = \exp(Y) = e^Y = \sum_{k=1}^{\infty} Y^k / k!$. The inverse operation $Y = \log(X)$ is called the matrix logarithm of X . In particular, if $Y \in \text{Sym}(p)$ and $Y = VLV'$ is an eigen-decomposition with $V \in O(p)$ and $L \in \text{Diag}(p)$, then it follows from the series expansion that

$$X = \exp(Y) = V \exp(L) V', \quad (39)$$

where $\exp(L)$ is a diagonal matrix containing the exponential of each of the diagonal elements of L . Since the diagonal elements of L are real, the diagonal elements of $\exp(L)$ are positive, implying that X is SPD. Given $X \in \text{Sym}^+(p)$, (39) can be inverted and the matrix log of X is uniquely given by

$$Y = \log(X) = V \log(\Lambda) V',$$

where $\log(\Lambda)$ is a diagonal matrix containing the logs of each of the diagonal elements of Λ .

Both $\exp(\cdot)$ and $\log(\cdot)$ are analytical functions and enjoy the useful properties with respect to similarity transformations and matrix inversion

$$\begin{aligned} \exp(AY A^{-1}) &= A \exp(Y) A^{-1}, & (e^Y)^{-1} &= e^{-Y} \\ \log(AX A^{-1}) &= A \log(X) A^{-1}, & \log(X^{-1}) &= -\log(X) \end{aligned} \quad (40)$$

for $A \in GL(p)$, the set of $p \times p$ real invertible matrices. In general, $e^{Y+Z} = e^Y e^Z$ and $\log(XW) = \log(X) + \log(W)$ unless Y and Z (or X and W) commute, i.e. they have the same eigenvectors. Instead, there is the Baker-Campbell-Hausdorff (BCH) formula

$$\log(e^Y e^Z) = Y + Z + \frac{1}{2}[Y, Z] + \frac{1}{12}[Y, [Y, Z]] + \frac{1}{12}[Z, [Z, Y]] + \frac{1}{24}[[Y, [Y, Z]], Z] + \dots$$

where $[Y, Z] = YZ - ZY$ is the matrix commutator. Notice that indeed $\log(e^Y e^Z) = Y + Z$ if Y and Z commute. In this paper we use an explicit polynomial expression without the commutators, also known as Goldberg's expansion (Goldberg, 1956). Using the method of Newman and Thompson (1987), the first several terms of this series are found to be

$$\begin{aligned} \log(e^Y e^Z) &= Y + Z + \frac{1}{2}(YZ - ZY) + \frac{1}{12}(YZ^2 + Y^2Z + ZY^2 + Z^2Y) \\ &\quad - \frac{1}{6}(YZY + ZYZ) \\ &\quad + \frac{1}{24}(Y^2Z^2 - Z^2Y^2) \\ &\quad - \frac{1}{12}(YZYZ - ZYZY) \\ &\quad - \frac{1}{720}(YZ^4 + Y^4Z + ZY^4 + Z^4Y) \\ &\quad + \frac{1}{180}(Y^2Z^3 + Y^3Z^2 + Z^2Y^3 + Z^3Y^2 + YZY^3 + YZ^3Y + Y^3ZY + ZYZ^3 + ZY^3Z + Z^3YZ) \\ &\quad - \frac{1}{120}(YZ^2Y^2 + Y^2ZY^2 + Y^2Z^2Y + ZY^2Z^2 + Z^2YZ^2 + Z^2Y^2Z) \\ &\quad - \frac{1}{120}(YZYZ^2 + YZY^2Z + YZ^2Z + YZ^2YZ + Y^2ZYZ + ZYZY^2 + ZYZ^2Y + ZY^2ZY + Z^2YZY) \\ &\quad + \frac{1}{30}(YZYZY + ZYZYZ) + \dots \end{aligned} \quad (41)$$

A.3 $\text{Sym}^+(p)$ as a Riemannian manifold with the Euclidean and log-Euclidean metrics

In the view of Section 2.2, $\text{Sym}^+(p)$ is seen as a Riemannian submanifold of $\text{Sym}(p)$ endowed with the Euclidean or Frobenius metric inherited from $\text{Sym}(p)$. Given two tangent vectors $Y, Z \in \text{Sym}(p)$, their Euclidean or Frobenius inner product and associated squared norm are defined as

$$\langle Y, Z \rangle = \text{tr}(YZ) = \text{vecd}(Y)' \text{vecd}(Z), \quad \|Y\|^2 = \text{tr}(Y^2), \quad (42)$$

and the geodesic distance between two points $X_1, X_2 \in \text{Sym}^+(p)$ is the Euclidean distance defined via (20).

A Riemannian manifold is called geodesically complete if at any point on the manifold, the Riemannian exponential map, which maps tangent vectors to the manifold along geodesics, is defined for all tangent vectors. This property depends on the metric that is defined on the manifold. When endowed with the Euclidean metric, $\text{Sym}^+(p)$ is not geodesically complete because the Riemannian exponential map from the tangent space $\text{Sym}(p)$ to $\text{Sym}^+(p)$ at any point, which for the Euclidean metric is equal to a displacement by a tangent vector, is only defined for those tangent vectors whose displacement falls within the boundaries of $\text{Sym}^+(p)$. As a consequence, the approximate normal distribution of the Euclidean average (5) is defined on $\text{Sym}(p)$ but not on $\text{Sym}^+(p)$.

One way of achieving geodesic completeness is via the global matrix log transformation (1), which effectively removes the SPD constraints by mapping each element of $\text{Sym}^+(p)$ to a unique element of $\text{Sym}(p)$ (Arsigny et al., 2006, 2007; Schwartzman, 2006). This is the view of Sections 3.1 and 3.2. According to this metric, the geodesic distance between two points $X_1, X_2 \in \text{Sym}^+(p)$ is the log-Euclidean distance defined via (21). Because of geodesic completeness, the approximate lognormal distribution of the log-Euclidean average (13) is well defined on $\text{Sym}^+(p)$.

As shown in Sections 3.1 and 3.2, the log transformation is simple enough that it permits obtaining results without the need of the Riemannian machinery.

A.4 $\text{Sym}^+(p)$ as a Riemannian manifold with the canonical metric

A more elaborate way of producing a geodesically complete manifold is via the so-called canonical or affine-invariant metric. For tangent vectors $Y, Z \in \text{Sym}(p)$ at $M \in \text{Sym}^+(p)$, their canonical or affine invariant inner product and associated squared norm are defined as

$$\langle Y, Z \rangle_M = \text{tr}(M^{-1}YM^{-1}Z), \quad \|Y\|_M^2 = \text{tr}\left[(M^{-1}Y)^2\right]. \quad (43)$$

Notice that at $M = I$, the $p \times p$ identity matrix, the canonical inner product (43) reduces to the Frobenius inner product (42). Various aspects of the geometry of $\text{Sym}^+(p)$ under the metric (43) have been described elsewhere (Skovgaard, 1984; Kobayashi and Nomizu, 1996; Lang, 1999; Moakher, 2005; Smith, 2005; Schwartzman, 2006; Lenglet et al., 2006; Zhu et al., 2009), which we summarize here for completeness and easier reference.

Define the linear group action of $GL(p)$ on $\text{Sym}^+(p)$

$$\phi(G, X) = GXG', \quad G \in GL(p), \quad X \in \text{Sym}^+(p), \quad (44)$$

which maps the SPD matrix X to the SPD matrix GXG' . The metric (43) is called affine-invariant because it has the isometric property of being invariant under the group action (44), that is

$$\langle\langle Y, Z \rangle\rangle_M = \langle\langle GYG', GZG' \rangle\rangle_{GMG'} \quad (45)$$

for any $G \in GL(p)$. In particular, if $G = M^{-1/2}$, then (45) becomes

$$\langle\langle Y, Z \rangle\rangle_M = \langle\langle M^{-1/2}YM^{-1/2}, M^{-1/2}ZM^{-1/2} \rangle\rangle_I = \text{tr} \left(M^{-1/2}YM^{-1/2} \cdot M^{-1/2}ZM^{-1/2} \right).$$

This says that the affine-invariant inner product can be computed by translating the tangent vectors to the tangent space at the identity and computing the standard Frobenius inner product there.

This device is useful for obtaining the Riemannian exponential and logarithmic maps corresponding to the affine-invariant metric. Since the Riemannian exponential map of a tangent vector $Y \in \text{Sym}(p)$ at the identity I is equal to the matrix exponential $\exp(Y) \in \text{Sym}^+(p)$, the Riemannian exponential map of a tangent vector Y at M onto $\text{Sym}^+(p)$ is obtained by mapping Y to the tangent space at the identity via $\phi(M^{-1/2}, \cdot)$, evaluating the Riemannian exponential map there, and mapping the result back again via $\phi(M^{1/2}, \cdot)$. The Riemannian exponential map is thus given by

$$X = \text{Exp}_M(Y) = M^{1/2} \exp \left(M^{-1/2}YM^{-1/2} \right) M^{1/2} = M \exp \left(M^{-1}Y \right), \quad (46)$$

where the second expression is obtained via (40). By (1), this map is one-to-one. The inverse map is the Riemannian logarithmic map of $X \in \text{Sym}^+(p)$ to the tangent vector $Y \in \text{Sym}(p)$ in the tangent space at M , and is given by

$$Y = \text{Log}_M(X) = M^{1/2} \log \left(M^{-1/2}XM^{-1/2} \right) M^{1/2} = M \log \left(M^{-1}X \right). \quad (47)$$

Because the Riemannian exponential map is bijective, there is a unique geodesic path joining any two points in $\text{Sym}^+(p)$. The geodesic distance between them may be defined uniquely as the length of that path, which by Gauss' lemma (DoCarmo, 1992, p. 69) is equal to the length of the projection via the Riemannian logarithmic map of one of the points onto the tangent space at the other. For any two points $M, X \in \text{Sym}^+(p)$, their geodesic distance according to the affine-invariant metric (43) is defined via (22).

The bijective property of the Riemannian exponential map makes $\text{Sym}^+(p)$ geodesically complete. Consequently, it is also complete as a metric space by the Hopf-Rinow theorem (DoCarmo, 1992, p. 147). The geodesic distance (22) is a proper metric and satisfies the properties of semi-positiveness, symmetry and the triangle inequality. In addition, it is invariant under similarity transformations, under the group action (44) and under inversion (Forstner and Moonen, 1999):

$$\begin{aligned} d_C(M, X) &= d_C(AMA^{-1}, AXA^{-1}), \quad \forall A \in GL(p) \\ d_C(M, X) &= d_C(GMG', GXG'), \quad \forall G \in GL(p) \\ d_C(M, X) &= d_C(M^{-1}, X^{-1}). \end{aligned} \quad (48)$$

These properties make $\text{Sym}^+(p)$ a Riemannian homogeneous space and a symmetric space in the sense of Cartan (Kobayashi and Nomizu, 1996, Ch. X-XI).

The curvature of $\text{Sym}^+(p)$ under the affine-invariant inner product is non-positive and has been derived explicitly by Lenglet et al. (2006). Intuitively, non-positive curvature means that geodesics “diverge” away from each other. This property guarantees the existence and uniqueness of the canonical geometric mean (Proposition 2).

B Proofs

B.1 Proof of Proposition 1

Proof 1—This result can be proven in at least two different ways. For the first proof, let $X = V\Lambda V'$ and $Y = VL V'$ be eigen-decompositions of X and Y so that $L = \log \Lambda$ with diagonal entries $l_j = \log \lambda_j$. Suppose the eigenvalues are all distinct. The Jacobian determinant of the eigenvalue decomposition $X \rightarrow (V, \Lambda)$ can be written as

$\mathcal{J}(X \rightarrow (V, \Lambda)) = g(V) \prod_{i < j} (\lambda_i - \lambda_j)$, where $g(V)$ is a function of the eigenvectors only (Mehta, 1991, p. 58). Similarly, $\mathcal{J}(Y \rightarrow (V, L)) = g(V) \prod_{i < j} (l_i - l_j)$. Finally, since $l_j = \log \lambda_j$, the transformation $L = \log \Lambda$ has Jacobian determinant $\mathcal{J}(L \rightarrow \Lambda) = |\Lambda|^{-1}$. Putting the three together,

$$\begin{aligned} \mathcal{J}(Y \rightarrow X) &= \mathcal{J}(Y \rightarrow (V, L)) \cdot \mathcal{J}((V, L) \rightarrow (V, \Lambda)) \cdot \mathcal{J}((V, \Lambda) \rightarrow X) \\ &= \left(g(V) \prod_{i < j} (l_i - l_j) \right) |\Lambda|^{-1} \left(g(V) \prod_{i < j} (\lambda_i - \lambda_j) \right)^{-1} \\ &= \frac{1}{\lambda_1 \dots \lambda_p} \prod_{i < j} \frac{\log \lambda_i - \log \lambda_j}{\lambda_i - \lambda_j}. \end{aligned}$$

The case $\lambda_i = \lambda_j$ is obtained by taking the limit of the fraction inside the product sign above as $\lambda_i \rightarrow \lambda_j$ equal to $d(\log \lambda_j)/d\lambda_j = 1/\lambda_j$.

Proof 2—An alternative proof is obtained by computing de Fréchet derivative of the matrix logarithm. For any $A \in \text{Sym}(p)$,

$$D(\log X)(A) = \lim_{t \rightarrow 0} \frac{\log(X+tA) - \log(X)}{t} = [g(\lambda_i, \lambda_j)] \circ A,$$

where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of X , $\circ$ denotes the Shur or entry-wise product of $p \times p$ matrices, and $[g(\lambda_i, \lambda_j)]$ is the Löwner matrix with (i, j) entries given by (15) (Bhatia and Holbrook, 2005). The Jacobian determinant is the product of the (non-repeated) entries of the Löwner matrix.

B.2 Proof of Theorem 1

B.2.1 Asymptotic normality—By Remark 2.2 of Bhattacharya and Patrangenaru (2005), Theorem 2.2 of Bhattacharya and Patrangenaru (2005) regarding the CLT for sample intrinsic means applies to $\text{Sym}^+(p)$ with the canonical inner product because it has no cut-locus and the supremum of its sectional curvatures is nonpositive. Note that the measure Q need not have compact support, as long as the second moment conditions in the statement of Theorem 1 are satisfied. Applying Theorem 2.2 of Bhattacharya and Patrangenaru (2005) implies that the conclusion of Theorem 2.1 of Bhattacharya and Patrangenaru (2005) regarding the CLT for sample Fréchet means holds, and we state it next.

Rather than working with the variables X_p it is convenient to work with the normalized variables $X_i^I = M^{-1/2} X_i M^{-1/2}$. By equivariance of the canonical geometric mean and average with respect to the group action of $GL(p)$ (Section 4.2) with $G = M^{-1/2}$, the population canonical geometric mean of the X_i^I 's is the identity matrix I (hence the superscript), and the canonical geometric average is $\bar{X}^I = M^{-1/2} \bar{X} M^{-1/2}$. To apply the CLT result above to the centered variables, let the distance ρ needed there be the geodesic distance d_c (22) and let the chart $\phi(\cdot)$ be the Riemannian logarithmic map $\text{Log}(\cdot) = \log(\cdot)$ (47) at I , which covers $\text{Sym}^+(p)$ by (1). For two tangent vectors $U, \Theta \in \text{Sym}(p)$ on the tangent space at I , the map ϕ induces the distance

$$\rho^\phi(U, \Theta) = \rho(\phi^{-1}(U), \phi^{-1}(\Theta)) = d_c(e^U, e^\Theta), \quad U, \Theta \in \text{Sym}(p) \quad (49)$$

Then the image measure Q^ϕ has Fréchet mean $M^\phi = \phi(M^I) = 0$ with respect to this distance.

Similarly, $\hat{M}_c^\phi = \phi\left(\bar{X}^I\right)$ is the Fréchet mean of the image empirical measure

$\hat{Q}_n^\phi = (1/n) \sum_{i=1}^n \delta(X_i^\phi)$, where

$$X_i^\phi = \phi(X_i^I) = \log(M^{-1/2} X_i M^{-1/2}) = Y_i$$

is assumed to have finite covariance. Further, from (49), define the half squared distance map

$$\Theta \rightarrow G(U; \Theta) = \frac{1}{2} (\rho^\phi)^2 (U, \Theta) = \frac{1}{2} d_c^2 (e^U, e^\Theta) \quad (50)$$

together with its gradient and Hessian at Θ

$$\Psi(U; \Theta) = \nabla_\Theta G(U; \Theta) \in \text{Sym}(p), \quad H(U; \Theta) = \partial_\Theta \text{vecd}[\Psi(U; \Theta)] \in GL(q). \quad (51)$$

Note the factor 1/2 in front of (50), missing in Bhattacharya and Patrangenaru (2005, p. 1229). Both the gradient and Hessian exist because the map (50) is at least twice differentiable. Then Theorem 2.1 of Bhattacharya and Patrangenaru (2005) states that

$$\sqrt{n} (\hat{M}_C^\phi - M_C^\phi) = \sqrt{n} \left[\log \left(M^{-1} \hat{X} M^{-1/2} \right) - 0 \right] \Rightarrow N(0, K^{-1} C K^{-1}).$$

where

$$C = \text{Cov} \left\{ \text{vecd} \left[\Psi(X_1^\phi; 0) \right] \right\} \in \text{Sym}^+(q), \quad K = E \left[H(X_1^\phi; 0) \right] \in GL(q), \quad (52)$$

provided that K is invertible. The required quantities C and K are computed in detail, next.

B.2.2 Computation of the covariance C —To compute the gradient in (51), we follow the method of Moakher (2005). Let $\Theta + tA$ with $t \in \mathbb{R}$ and $A \in \text{Sym}(p)$ be a straight line in the tangent space at I . From (50), define the scalar function

$$h(t) = G(U; \Theta + tA) = \frac{1}{2} d_c^2 (e^U, e^{\Theta + tA}) = \frac{1}{2} \text{tr} \left[\log^2 (e^{-U} e^{\Theta + tA}) \right] \quad (53)$$

where the last expression was obtained using (46) and (22). Then the gradient in (51) is the unique tangent vector $\Psi(U; \Theta)$ such that

$$\left. \frac{dh(t)}{dt} \right|_{t=0} = \langle \Psi(U; \Theta), A \rangle_I = \text{tr}(\Psi(U; \Theta) A), \quad (54)$$

where the Riemannian inner product (43) on the tangent space at I reduces to the Frobenius inner product (42).

Using the chain rule formula (Moakher, 2005, Prop. 2.1)

$$\frac{d}{dt} \text{tr} \left[\log^2 X(t) \right] = 2 \text{tr} \left[\log X(t) X^{-1}(t) \frac{d}{dt} X(t) \right]$$

in (53) with $X(t) = e^{-U}e^{\Theta+tA}$ gives that the derivative of $h(t)$ at 0 is

$$\left. \frac{dh(t)}{dt} \right|_{t=0} = \text{tr} \left[\log(e^{-U}e^{\Theta}) \cdot e^{-\Theta} \left. \frac{d}{dt} e^{\Theta+tA} \right|_{t=0} \right]. \quad (55)$$

The first term inside the brackets has a polynomial expansion given by Goldberg's formula (41). Keeping terms only up to first order in Θ ,

$$\begin{aligned} \log(e^{-U}e^{\Theta}) &= -U + \Theta + \frac{1}{2}(-U\Theta + \Theta U) + \frac{1}{12}(U^2\Theta + \Theta U^2) \\ &\quad - \frac{1}{6}U\Theta U \\ &\quad - \frac{1}{720}(U^4\Theta + \Theta U^4) \\ &\quad + \frac{1}{180}(U\Theta U^3 + U^3\Theta U) \\ &\quad - \frac{1}{120}U^2\Theta U^2 + \dots \end{aligned}$$

Similarly, keeping terms only up to first order in Θ and t in the second term in the brackets in (55),

$$e^{-\Theta} \left. \frac{d}{dt} e^{\Theta+tA} \right|_{t=0} = (I - \Theta + \dots) \frac{d}{dt} \left[I + \Theta + tA + \frac{1}{2}(\Theta + tA)^2 + \dots \right]_{t=0} = A + \frac{1}{2}(A\Theta - \Theta A) + \dots$$

Replacing in (55),

$$\left. \frac{dh(t)}{dt} \right|_{t=0} = \text{tr} \left\{ \left[-U + \Theta + \frac{1}{12}(U^2\Theta - 2U\Theta U + \Theta U^2) + \frac{1}{720}(U^4\Theta - 4U^3\Theta U + 6U^2\Theta U^2 - 4U\Theta U^3 + \Theta U^4) + \dots \right] A \right\}.$$

Therefore, comparing with (54), we have that the desired gradient is

$$\Psi(U; \Theta) = -U + \Theta + \frac{1}{12}(U^2\Theta - 2U\Theta U + \Theta U^2) + \frac{1}{720}(U^4\Theta - 4U^3\Theta U + 6U^2\Theta U^2 - 4U\Theta U^3 + \Theta U^4) + \dots \quad (56)$$

which is indeed in $\text{Sym}(p)$. In particular, $\Psi(X_1^\psi; 0) = -X_1^\phi = -Y_1$, with covariance $C = \text{Cov}(\text{vecd}(Y_1)) = \Sigma$.

B.2.3 Computation of the expected Hessian K —To obtain the Hessian in (51), we first vectorize the gradient (56) as

$$vecd[\Psi(U; \Theta)] = -vecd(U) + vecd(\Theta) + D' \left[\frac{1}{2}(U \otimes I - I \otimes U)^2 + \frac{1}{720}(U \otimes I - I \otimes U)^4 + \dots \right] D vecd(\Theta)$$

where we have used property (38) and the property that $vec(AX - B) = (B' \otimes A) vec(X)$.
Thus

$$H(U; \Theta) = \frac{\partial vecd[\Psi(U; \Theta)]}{\partial [vecd(\Theta)]'} = I + D' \left[\frac{1}{12}(U \otimes U)^2 + \frac{1}{720}(U \otimes U)^4 + \dots \right] D$$

Finally, by (52) with $X_1^\phi = Y_1$, the desired matrix K is as given in (28). Note that K is SPD and thus invertible.

References

- Alexander AL, Hasan K, Kindlmann G, Parker DL, Tsuruda JS. A geometric analysis of diffusion tensor measurements of the human brain. *Magn Reson Med*. 2000; 44(2):283-291.
- Anderson, TW. An Introduction to Multivariate Statistical Analysis. third edition. John Wiley & Sons Ltd.; Hoboken, New Jersey: 2003.
- Arsigny V, Fillard P, Pennec X, Ayache N. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magn Reson Med*. 2006; 56(2):411-421. [PubMed: 16788917]
- Arsigny V, Fillard P, Pennec X, Ayache N. Geometric Means in a Novel Space Structure of Symmetric Positive Definite Matrices. *SIAM J Matrix Anal Appl*. 2007; 29(1):328-347.
- Ashburner J, Friston KJ. Nonlinear spatial normalization using basis functions. *Hum Brain Map*. 1999; 7:254-266.
- Bammer R, Auer M, Keeling SL, Agustin M, Stables LA, Prokesch RW, Stollberger R, Moseley ME, Fazekas F. Diffusion tensor imaging using single-shot SENSE-EPI. *Magn Reson Med*. 2002; 48:128-136. [PubMed: 12111940]
- Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J Magn Reson B*. 1996; 111:209-219. [PubMed: 8661285]
- Batchelor PG, Moakher M, Atkinson D, Calamante F, Connelly A. A rigorous framework for diffusion tensor calculus. *Magn Reson Med*. 2005; 53:221-225. [PubMed: 15690523]
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B*. 1995; 57(1):289-300.
- Bhatia, R. Positive Definite Matrices. Princeton University Press; 2007.
- Bhatia R, Holbrook J. Riemannian geometry and matrix geometric means. *Linear Algebra and its Applications*. 2006; 413:594-618.
- Bhattacharya A, Bhattacharya RN. Statistics on Riemannian manifolds: asymptotic distribution and curvature. *Proceedings of the American Mathematical Society*. 2008; 33(3):1225-1259.
- Bhattacharya RN, Elligson L, Liu X, Patrangenaru V, Crane M. Extrinsic analysis on manifolds is computationally faster than intrinsic analysis, with applications to quality control by machine vision. *Appl. Stochastic Models Bus. Ind.* 2012; 28:222-235.
- Bhattacharya RN, Patrangenaru V. Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann Statist*. 2003; 31(1):1-29.
- Bhattacharya RN, Patrangenaru V. Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann Statist*. 2005; 33(3):1225-1259.
- Calvo M, Oller JM. An explicit solution of information geodesic equations for the multivariate normal model. *Statistics and Decisions*. 1991; 9:119-138.

- Chiu TYM, Leonard T, Tsui KW. The matrix-logarithmic covariance model. *J Amer Statist Assoc.* 1996; 91(433):198–210.
- DoCarmo, MP. *Riemannian Geometry*. Birkhauser; Boston: 1992.
- Dougherty RF, Ben-Shachar M, Deutsch GK, Hernandez A, Fox GR, Wandell BA. Temporal-callosal pathway diffusivity predicts phonological skills in children. *PNAS.* 2007; 104(20):8556–8561. [PubMed: 17483487]
- Dougherty RF, Ben-Shachar M, Deutsch GK, Potanina P, Wandell BA. Occipital-callosal pathways in children: Validation and atlas development. *Ann N Y Acad Sci.* 2005; 1064:98–112. [PubMed: 16394151]
- Dryden IL, Koloydenko A, Zhou D. Non-Euclidean statistics for covariance matrices, with applicationns to diffusion tensor imaging. *Ann of Appl Statist.* 2009; 3(3):1102–1123.
- Ellingson, L.; Groisser, D.; Osborne, D.; Patrangenaru, V.; Schwartzman, A. Nonparametric bootstrap of sample means of positive definite matrices with an application to diffusion tensor imaging data analysis.; Technical Report. 2012. p. M1005<http://www.stat.fsu.edu/techreports.php>
- Fillard P, Pennec X, Arsigny V, Ayache N. Clinical DT-MRI Estimation, Smoothing, and Fiber Tracking With Log-Euclidean Metrics. *IEEE-Transactions in Medical Imaging.* 2007; 26(11): 1472–1482.
- Fletcher PT, Joshi S. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing.* 2007; 87(2):250–262.
- Forstner, W.; Moonen, B. A metric for covariance matrices.. In: Schwarze, und Volker S.; F. K., editors. *Quo vadis geodesia...?*, Festschrift for Erik W. Grafarend on the occasion of his 60th birthday, number 1999.6 in Technical Reports Department of Geodesy and Geoinformatics. Stuttgart University; 1999. p. 113-128.
- Fréchet M. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann Inst H Poincaré.* 1948; 10:215–310.
- Goldberg K. The formal power series for $\log(e^X e^Y)$. *Duke Mathematical Journal.* 1956; 23:13–21.
- Groisser D. Newton's method, zeroes of vector fields, and the Riemannian center of mass. *Adv. Appl. Math.* 2004; 33:95–135.
- Gupta, AK.; Nagar, DK. *Matrix Variate Distributions*. Chapman & Hall/CRC; Boca Raton, Florida: 2000.
- Henderson HV, Searle SR. Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *Can. J. Stat.* 1979; 7:6581.
- Johnson, NL.; Kotz, S. *Continuous Univariate Distributions-1*. John Wiley & Sons; New York: 1970.
- Joshi AA, Shattuck DW, Thompson PM, Leahy RM. Surface-constrained volumetric brain registration using harmonic mappings. *IEEE-Transactions in Medical Imaging.* 2007; 26(12):1657–1669.
- Kobayashi, S.; Nomizu, K. *Fundamentals of Differential Geometry - Volume II*. John Wiley & Sons, Inc.; New York: 1996.
- Lang, S. *Fundamentals of Differential Geometry*. Springer-Verlag; New York: 1999.
- Lawson JD, Lim Y. The geometric mean, matrices, metrics, and more. *The American Mathematical Monthly.* Nov.2001 :797–812.
- Lawson JD, Lim Y. Weighted means and Karcher equations of positive operators. *PNAS.* 2013; 110(39):1562615632.
- LeBihan D, Mangin J-F, Poupon C, Clark CA, Pappata S, Molko N, Chabriat H. Diffusion tensor imaging: Concepts and applications. *J Magn Reson Imaging.* 2001; 13:534–546. [PubMed: 11276097]
- Lenglet C, Rousson M, Deriche R, Faugeras O. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *J of Math Imaging and Vision.* 2006; 25(3):423–444.
- Leonard T, Hsu JSJ. Bayesian inference for a covariance matrix. *Ann Statist.* 1992; 20(4):1669–1696.
- Leporé N, Brun C, Chou Y-Y, Chiang M-C, Dutton RA, Hayashi KM, Lueders E, Lopez OL, Aizenstein HJ, Toga AW, Becker JT, Thompson PM. Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. *IEEE-Transactions in Medical Imaging.* 2008; 27(1):129–141.

- Magnus JR, Neudecker H. The elimination matrix: some lemmas and applications. *SIAM J. Algebr. Discrete Methods*. 1980; 1:422-449.
- Mardia, KV.; Kent, JT.; Bibby, JM. *Multivariate Analysis*. Academic Press; San Diego, California: 1979.
- Mehrabadi MM, Cowin SC, Jaric J. Six-dimensional Orthogonal Tensor Representation of the Rotation about an Axis in Three Dimensions. *Int J Solids Structures*. 1995; 32:439–449. 3/4.
- Mehta, ML. *Random Matrices*. 2nd edition. Academic Press; San Diego, California: 1991.
- Moakher M. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Matrix Anal Appl*. 2005; 26(3):735–747.
- Moakher, M.; Batchelor, PG. *Visualization and Processing of Tensor Fields*. Springer; 2006.
- Symmetric Positive-Definite Matrices: From Geometry to Applications and Visualization.. Ch. 17
- Moakher M, Zérai. The Riemannian Geometry of the Space of Positive-Definite Matrices and Its Application to the Regularization of Positive-Definite Matrix-Valued Data. *J Math Imaging Vis*. 2011; 40:171-187.
- Muirhead, RJ. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons Ltd.; Hoboken, New Jersey: 1982.
- Newman M, Thompson RC. Numerical values of Goldberg's coefficients in the series for $\log(e^X e^Y)$. *Mathematics of computation*. 1987; 48(177):265–271.
- Oller JM, Corcuera JM. Intrinsic analysis of statistical estimation. *Ann Statist*. 1995; 23(5):1562-1581.
- Osborne D, Patrangenaru V, Ellingson L, Groisser D, Schwartzman A. Non-parametric two-sample tests on homogeneous Riemannian manifolds, Cholesky decompositions and diffusion tensor image analysis. *J Multivariate Anal*. 2013; 119:163–175.
- Pasternak O, Sochen N, Basser PJ. The effect of metric selection on the analysis of diffusion tensor MRI data. *Neuroimage*. 2010; 49(3):2190–2204. [PubMed: 19879947]
- Pennec X, Fillard P, Ayache N. A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*. 2006; 66(1):41-66.
- Pennec X. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J Math Imaging Vis*. 2006; 25:127-154.
- Schott, JR. *Matrix Analysis for Statistics*. 2nd edition. John Wiley & Sons Ltd.; Hoboken, New Jersey: 2005.
- Schwartzman, A. PhD thesis. Stanford University; 2006. Random ellipsoids and false discovery rates: statistics for diffusion tensor imaging data..
- Schwartzman A, Dougherty RF, Taylor JE. Group comparison of eigenvalues and eigenvectors of diffusion tensors. *J Amer Statist Assoc*. 2010; 105(490):588–599.
- Schwartzman A, Mascarenhas W, Taylor JE. Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. *Ann Statist*. 2008; 36(6):2886–2919.
- Shurbet GL, Lewis TO, Boullion TL. Quadratic Matrix Equations. *The Ohio Journal of Science*. 1974; 74(5):273–277.
- Skovgaard LT. A Riemannian Geometry of the Multivariate Normal Model. *Scandinavian Journal of Statistics*. 1984; 11(4):211–223.
- Smith S. Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Trans Signal Proc*. 2005; 53(5):1610–1630.
- Terras A. Asymptotics of Special Functions and the Central Limit Theorem on the Space P_n of Positive $n \times n$ Matrices. *J Multivariate Anal*. 1987; 23:13–36.
- Whitcher B, Wisco JJ, Hadjikhani N, Tuch DS. Statistical group comparison of diffusion tensors via multivariate hypothesis testing. *Magn Reson Med*. 2007; 57:1065–1074. [PubMed: 17534902]
- Yao Y. An Approximate Degrees of Freedom Solution to the Multivariate Behrens Fisher Problem. *Biometrika*. 1965; 52:139-147.
- Zhu H, Chen Y, Ibrahim JG, Li Y, Hall C, Lin W. Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *J Amer Statist Assoc*. 2009; 104(487): 1203–1212.

Figure 1.

A 2×2 matrix X with diagonals a, b and off-diagonal c is SPD if and only if $a > 0$, $b > 0$ and

$ab - c^2 > 0$. The set of valid triplets $\text{vecd}(X) = (a, b, \sqrt{2}c)'$ (see also Appendix A.1) is an open cone in $\mathbb{R}^3$ with circular cross-sections. The diagonal matrices $X_1 = \text{Diag}(0.9, 0.1)$ and $X_2 = \text{Diag}(0.1, 0.9)$ are represented by the triplets $\text{vecd}(X_1) = (0.9, 0.1, 0)'$ and $\text{vecd}(X_2) = (0.1, 0.9, 0)'$ respectively. Because this set is convex, interpolation between any two points is permitted, and the Euclidean average $\bar{X} = \text{Diag}(0.5, 0.5)$ is SPD. Extrapolation, however, might result in matrices that are not SPD: the green straight line connecting X_1 and X_2 extends beyond the boundaries of the cone. The yellow hyperbola connecting X_1 and X_2 is the matrix exponential of the straight line between $\log X_1$ and $\log X_2$, always inside the cone by construction. The midpoint along this line is the log-Euclidean average (10) equal to $\hat{X} = \text{Diag}(0.3, 0.3)$. In this case, it also coincides with the canonical geometric average defined via (25).

Figure 2.

(a) 30 samples (gray) generated from a lognormal distribution $X \sim LN(M, \Sigma)$ with parameters $M = \text{Diag}(4, 1)$ and $\Sigma = \text{Diag}(0.5, 0.5, 0)$ represented by ellipses via (2). Notice the variation around the mean ellipse M (red) with semiaxes $\sqrt{4}=2$ and $\sqrt{1}=1$. Superimposed are the log-Euclidean average (inner blue) and two extreme points of the 95% CR (outer blue). (b) Same as in (a) displayed in the coordinates of Figure 1, where each ellipse in (a) corresponds to a point of the same color inside the cone. The 95% CR is represented by the cyan mesh.

Figure 3.

(a) Density of $y = \text{vecd}(Y) \sim N(\text{vecd}(\log M), \Sigma)$ with parameters (17) displayed in two planes: $y_3 = 0$ (left column) and $y_1 = y_2$ (right column). The density contours increase in value for darker colors. (b) Density of $x = \text{vecd}(X)$ with $X = \exp(Y)$ displayed in two planes: $x_3 = 0$ (left column) and $x_1 = x_2$ (right column).

Figure 4.

(a) Density of $y = \text{vecd}(Y) \sim \mathcal{N}(0, \Sigma)$ with parameters (17) displayed in two planes: $y_3 = 0$ (left column) and $y_1 = y_2$ (right column). The density contours increase in height for darker colors. (b) Density of $x = \text{vecd}(X)$ with $X = \exp(M^{1/2} Y M^{1/2})$ displayed in two planes: $x_3 = 0$ (left column) and $x_1 = x_2$ (right column).

Figure 5.

Simulated p-value distributions under equality of means (left) and under the largest mean separation found in the DTI data (right) for 3000 repetitions. The data were generated from a canonical lognormal distribution, while the tests were performed for the three different geometries.

Figure 6. Pooled voxelwise DT averages. Image intensity is proportional to FA in the range $[0, 0.6]$. Colors indicate coordinate directions of the PDD: right-left (red), anterior-posterior (green) and superior-inferior (blue). Mixed colors represent directions that are oblique to the coordinate axes.

Figure 7.

(a) FA differences between: log-Euclidean and Euclidean average (left); log-Euclidean and canonical (right). Image intensity scale is the same for both panels. (b) Angle between: PDD of log-Euclidean and Euclidean average (left); PDD of log-Euclidean and canonical (right). Image intensity scale in degrees is the same in both panels.

Figure 8. Ellipsoidal rendering of the DT pooled averages (red) and two extreme points of the corresponding 95% CR (blue) at voxel with largest FA difference between average types. Colored axes correspond to the coordinate directions described in Figure 6.

Figure 9. Maps of two-sample p-values in $\log_{10}$ scale. The p-values produced by the three types of averages are indistinguishable.

Figure 10.
Two-sample p-value distributions in the DTI data.