Bernoulli **31**(2), 2025, 1552–1578
https://doi.org/10.3150/24-BEJ1783

Averaging symmetric positive-definite matrices on the space of eigen-decompositions

SUNGKYU JUNG[1,a], BRIAN ROOKS[2,b], DAVID GROISSER[3,c] and ARMIN SCHWARTZMAN[4,d]

[1]*Department of Statistics and Institute for Data Innovation in Science, Seoul National University, Gwanak-gu, Seoul 08826, South Korea,* [a]*sungkyu@snu.ac.kr*
[2]*Statistics & Data Corporation, Tempe, AZ 85288, USA,* [b]*btrooks88@gmail.com*
[3]*Department of Mathematics, University of Florida, Gainesville, FL 32611, USA,* [c]*groisser@ufl.edu*
[4]*Halıcıoğlu Data Science Institute and Division of Biostatistics, University of California, San Diego, La Jolla, CA 92093, USA,* [d]*armins@ucsd.edu*

We study extensions of Fréchet means for random objects in the space $\mathrm{Sym}^+(p)$ of $p \times p$ symmetric positive-definite matrices using the scaling-rotation geometric framework introduced by Jung et al. [*SIAM J. Matrix. Anal. Appl.* **36** (2015) 1180-1201]. The scaling-rotation framework is designed to enjoy a clearer interpretation of the changes in random ellipsoids in terms of scaling and rotation. In this work, we formally define the *scaling-rotation (SR) mean set* to be the set of Fréchet means in $\mathrm{Sym}^+(p)$ with respect to the scaling-rotation distance. Since computing such means requires a difficult optimization, we also define the *partial scaling-rotation (PSR) mean set* lying in the space of eigen-decompositions as a proxy for the SR mean set. The PSR mean set is easier to compute and its projection to $\mathrm{Sym}^+(p)$ often coincides with SR mean set. Minimal conditions are required to ensure that the mean sets are non-empty. Because eigen-decompositions are never unique, neither are PSR means, but we give sufficient conditions for the sample PSR mean to be unique up to the action of a certain finite group. We also establish strong consistency of the sample PSR means as estimators of the population PSR mean set, and a central limit theorem. In an application to multivariate tensor-based morphometry, we demonstrate that a two-group test using the proposed PSR means can have greater power than the two-group test using the usual affine-invariant geometric framework for symmetric positive-definite matrices.

Keywords: Central limit theorem; manifolds; scaling-rotation distance; strong consistency

1. Introduction

Recently, much work has been done to advance the statistical analysis of random symmetric positive-definite (SPD) matrices. Applications in which data arise as SPD matrices include the analysis of diffusion tensor imaging (DTI) data (Alexander, 2005, Batchelor et al., 2005), multivariate tensor-based morphometry (TBM) (Lepore et al., 2008, Paquette et al., 2017), and tensor computing (Pennec, Fillard and Ayache, 2006). In this paper, we consider the setting in which we have a random sample of SPD matrices and wish to estimate a population mean.

Location estimation is an important first step in the development of many statistical techniques. For applications in which data are SPD matrices, these techniques include two-sample hypothesis testing (Schwartzman, Dougherty and Taylor, 2010) for comparing average brain scans from two groups of interest, principal geodesic analysis (Fletcher et al., 2004) for visualizing major modes of variation in a sample of SPD matrices, and weighted mean estimation, which has useful applications in diffusion tensor processing, including fiber tracking, smoothing, and interpolation (Batchelor et al., 2005, Carmichael et al., 2013).

One of the challenges of developing methods for analyzing SPD-valued data is that the positive-definiteness constraint precludes $\mathrm{Sym}^+(p)$, the space of $p \times p$ SPD matrices, from being a vector sub-

1350-7265

space of Sym(p), the space of all symmetric $p \times p$ matrices. This can be easily visualized for $p = 2$; the free coordinates (two diagonal elements and one upper off-diagonal element) of all 2×2 SPD matrices in Sym(2) $\cong \mathbb{R}^3$ constitutes an open convex cone. Hence, conventional estimation or inferential techniques developed for data that vary freely over Euclidean space may not be appropriate for the statistical analysis of SPD matrices. With this in mind, many location estimation frameworks for $\mathrm{Sym}^+(p)$ have been developed in recent years, including the log-Euclidean framework (Arsigny et al., 2006/07), affine-invariant framework (Fletcher et al., 2004, Pennec, Fillard and Ayache, 2006), log-Cholesky framework (Lin, 2019), and Procrustes framework (Dryden, Koloydenko and Zhou, 2009, Masarotto, Panaretos and Zemel, 2019); see Feragen and Fuster (2017) for other examples. Given a sample of SPD matrices, most of these estimation methods amount to transforming the SPD-valued observations, averaging in the space of the transformed observations, and then mapping the mean of the transformed data into $\mathrm{Sym}^+(p)$. For example, the log-Euclidean method maps each observation into Sym(p) via the matrix logarithm, computes the sample mean of the transformed observations, and then maps that mean into $\mathrm{Sym}^+(p)$ via the matrix exponential function, while the Procrustes size-and-shape method begins with averaging the Cholesky square roots of observations, and then maps the average $\hat{L}$ to $\mathrm{Sym}^+(p)$ as $\hat{\Sigma} = \hat{L}\hat{L}^T$, where A^T denotes the transpose of a matrix A.

While these geometric frameworks account for the positive-definiteness constraint of $\mathrm{Sym}^+(p)$, it is not clear which, if any, of the log-Euclidean, affine-invariant, or Procrustes size-and-shape frameworks is most "natural" for describing deformations of SPD matrices. Motivated by the analysis of DTI data, a setting in which observations are SPD matrices represented as ellipsoids in $\mathbb{R}^3$, Jung, Schwartzman and Groisser (2015) developed a different framework, called the scaling-rotation (SR) framework for $\mathrm{Sym}^+(p)$. Under this framework, the distance between SPD matrices X and Y is defined as the minimal amount of rotation of axes and scaling of axis lengths necessary to deform the ellipsoid associated with X into the ellipsoid associated with Y. For this, an SPD matrix X is decomposed into eigenvectors and eigenvalues, which respectively stand for rotations and scalings. The SR framework yields interpolation curves that have desirable properties, including constant rate of rotation and log-linear scaling of eigenvalues, and it is the only geometric framework (compared to the aforementioned frameworks) to produce both pure-scaling interpolation curves and pure-rotation curves when the endpoints differ by pure scaling or pure rotation. While interpolation approaches similar to the SR framework can be found in Wang et al. (2014) and Collard et al. (2014), only the SR framework addresses the non-uniqueness of eigen-decompositions (Groisser, Jung and Schwartzman, 2017, 2021). See Feragen and Fuster (2017) and Feragen and Nye (2020) for a comparison of the SR framework with other geometric frameworks for SPD matrices.

A major complication in developing statistical procedures using the SR framework is that eigen-decompositions are not unique. For example, an SPD matrix $X = \mathrm{diag}(8,3) = \begin{pmatrix} 8 & 0 \\ 0 & 3 \end{pmatrix}$ can be eigen-decomposed into either

$$X = U_1 D_1 U_1^T, \quad U_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; D_1 = \begin{pmatrix} 8 & 0 \\ 0 & 3 \end{pmatrix},$$

or

$$X = U_2 D_2 U_2^T, \quad U_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \; D_2 = \begin{pmatrix} 3 & 0 \\ 0 & 8 \end{pmatrix}.$$

(There are in fact 4 distinct eigen-decompositions for diag(8,3), if the eigenvector matrices are required to be orthogonal matrices of positive determinant.) Write (U_X, D_X) for an eigen-decomposition (a pair of eigenvector and eigenvalue matrices) of an SPD matrix X, and let $\mathcal{F}$ be the eigen-composition map, e.g., $\mathcal{F}(U_X, D_X) = U_X D_X U_X^T = X$ (see Definition 2.1). The SR framework defines the "distance" between $X, Y \in \mathrm{Sym}^+(p)$ to be $d_{\mathcal{SR}}(X,Y) := \inf d_M((U_X, D_X),(U_Y, D_Y))$, where the infimum is taken over all possible eigen-decompositions of both X and Y, and d_M is the (geodesic) distance function on the space $M(p)$ of eigen-decompositions (see Definition 2.2). $\mathrm{Sym}^+(p)$ is a stratified space; the stratum to which $X \in \mathrm{Sym}^+(p)$ belongs is determined by the eigenvalue-multiplicity type of X, or equivalently by the topological structure of the fiber $\mathcal{F}^{-1}(X)$ (the set of all eigen-decompositions corresponding to $X \in \mathrm{Sym}^+(p)$); see Section 2.3. The scaling-rotation distance $d_{\mathcal{SR}}$ fails to be a true metric on $\mathrm{Sym}^+(p)$, and is difficult to compute because the set we minimize over in the definition of $d_{\mathcal{SR}}(X,Y)$ is a pair of these fibers (whose topology varies with the strata of X and Y). In fact, $\mathrm{Sym}^+(p)$ equipped with $d_{\mathcal{SR}}$ is not a Riemannian manifold; it is only the eigen-decomposition space $M(p)$ that we are endowing with a Riemannian metric. With these complications in mind, the goal of this paper is to establish location-estimation methods using the SR framework as a foundation for future methods that will inherit the interpretability of the framework.

If one of the well-established geometric frameworks, such as the affine-invariant or log-Cholesky frameworks, is used, then $\mathrm{Sym}^+(p)$ is understood as a Riemannian manifold with a Riemannian metric tensor defined on the tangent bundle. The Riemannian metric gives rise to a distance function, say d, and $(\mathrm{Sym}^+(p), d)$ is a metric space. For these metric spaces, the Fréchet mean (Fréchet, 1948) is a natural candidate for a location parameter, and conditions that guarantee uniqueness of Fréchet means, convergence of empirical Fréchet means to the population counterpart, and central-limit-theorem type results, are well-known (*cf.* Afsari, 2011, Bhattacharya and Lin, 2017, Bhattacharya and Patrangenaru, 2003, 2005, Eltzner et al., 2021, Huckemann, 2011a,b, Schötz, 2022).

But in the SR framework, since $d_{\mathcal{SR}}$ is not a true metric on $\mathrm{Sym}^+(p)$, many of the theoretical properties of Fréchet means (if they are defined) are no longer guaranteed. Moreover, on a practical side, computing a scaling-rotation (SR) mean, defined as a minimizer over the sum of squared SR distances to observations, requires discrete optimization in general and is thus challenging to implement. As a proxy for the SR mean, we define a *partial scaling-rotation (PSR) mean* on the space of eigen-decompositions; for a finite sample $X_1, \dots, X_n \in \mathrm{Sym}^+(p)$, the PSR mean set is the set of minimizers

$$\underset{(U,D)}{\mathrm{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left\{ \inf_{(U_X,D_X) \in \mathcal{F}^{-1}(X_i)} d_M((U_X,D_X),(U,D)) \right\}^2. \tag{1.1}$$

See Section 3 for precise definitions and an iterative algorithm for computing a sample PSR mean. The PSR means can be thought of as a special case of generalized Fréchet means, proposed in Huckemann (2011b) and studied in Huckemann (2011a), Huckemann and Eltzner (2021), Schötz (2019, 2022). The PSR means can be mapped to $\mathrm{Sym}^+(p)$ (via the eigen-composition map $\mathcal{F}$), and we establish some sufficient conditions under which the PSR means are *equivalent* to the SR mean. These conditions are related to the strata of $\mathrm{Sym}^+(p)$ in which the sample and means are located.

Another artifact caused by the stratification of $\mathrm{Sym}^+(p)$ is that the distance function $d_{\mathcal{SR}}$ is not continuous on $\mathrm{Sym}^+(p)$, and in principle we do not know whether an SR mean is well-defined. We show that the distance function $d_{\mathcal{SR}}$, the cost function appeared in (1.1) for the PSR means, and their squares are *lower semicontinuous*, and thus are measurable, which guarantees that both the SR and PSR mean sets are well-defined. We also show that SR and PSR mean sets exist, under mild assumptions.

PSR means are never unique, due to the fact that eigen-decompositions are not unique. In the best case, there are $2^{p-1}p!$ elements in the PSR mean set for a $\mathrm{Sym}^+(p)$-valued sample, corresponding to the number of distinct eigen-decompositions of any SPD matrix with no repeated eigenvalues. As a result, if a PSR mean set $E_n^{(\mathcal{PSR})}$ consists of exactly $2^{p-1}p!$ elements, then the corresponding $\mathrm{Sym}^+(p)$-valued mean, $\mathcal{F}(E_n^{(\mathcal{PSR})})$ consists of a single element, and we may say that $\mathcal{F}(E_n^{(\mathcal{PSR})})$ is unique. A sufficient condition to ensure such uniqueness will be given in Section 4.3 in terms of data-support diameter.

We also show that with only a finite-variance condition the sample PSR mean set is consistent with the population PSR mean set, in the sense of Bhattacharya and Patrangenaru (2003), following the now standard technique laid out in Huckemann (2011b) (with modifications required to the fact that the cost function in (1.1) is not continuous). With additional conditions, needed to ensure the equivalence between PSR mean sets and the SR mean, imposed, we conclude that the sample SR mean set is consistent with the (unique) SR mean. A type of central limit theorem for the PSR mean is also developed, in which the limiting normal distribution is defined on a tangent space of the space of eigen-decompositions. See Section 4 for theoretical properties of (partial) SR means, including existence, uniqueness, and asymptotic results. Although these properties are developed to cope with the unique challenges (e.g. non-uniqueness of eigen-decompositions and the resulting stratification) coming from using the SR framework, we believe the course of our technical development will be instructive for developing statistics in other stratified Riemannian spaces.

Numerical results demonstrate the subtle difference between the SR mean and the PSR mean, and the advantage of (partial) SR means over other means defined via other geometric frameworks. The potential advantage of the SR framework with PSR means is further demonstrated in an application to multivariate TBM for testing the shape difference in lateral ventricular structure in the brains of pre-term and full-term infants, using data from Paquette et al. (2017). In particular, an approximate bootstrap test based on PSR means is found to be more powerful than that based on the affine-invariant means of Pennec, Fillard and Ayache (2006). We conclude with practical advice on the analysis of SPD matrices and a discussion of potential future directions of research.

We organize the rest of this article as follows. The scaling-rotation geometric framework is reviewed in Section 2. In Section 3, we introduce the novel partial scaling-rotation (PSR) means and an estimation algorithm. Section 4 is devoted to theoretical properties of the PSR means, including sufficient conditions for existence and uniqueness, as well as consistency and a central limit theorem. The PSR means are numerically and visually demonstrated in Section 5, in which we also provide comparisons to other geometric frameworks for SPD matrices. We summarize our conclusions in Section 6. Technical details, proofs, and additional lemmas that may be useful in other contexts, are contained in the Supplementary Material (Jung et al., 2024).

2. The scaling-rotation framework

In this section we provide a brief overview of the scaling-rotation framework (Groisser, Jung and Schwartzman, 2017, 2021, Jung, Schwartzman and Groisser, 2015) for analyzing SPD-valued data. The motivation for the scaling-rotation framework is intuitive: Any $X \in \mathrm{Sym}^+(p)$ can be identified with the ellipsoid with surface coordinates $\{y \in \mathbb{R}^p : y^T X^{-1} y = 1\}$, so a measure of distance between X and Y can be defined as a suitable combination of the minimum amount of rotation of axes and stretching or shrinking of axes needed to deform the ellipsoid corresponding to X into the ellipsoid associated with Y. Since the semi-axes and squared semi-axis lengths of the ellipsoid associated with an SPD matrix are its eigenvectors and eigenvalues, respectively, this *scaling-rotation distance* is computed on the space of eigen-decompositions.

2.1. Geometry of the eigen-decomposition space

Recall that any $X \in \mathrm{Sym}^+(p)$ has an eigen-decomposition $X = UDU^T$, where $U \in SO(p)$, the space of $p \times p$ rotation matrices, and $D \in \mathrm{Diag}^+(p)$, the space of $p \times p$ diagonal matrices possessing positive diagonal entries. We denote the space of eigen-decompositions as $M(p) := SO(p) \times \mathrm{Diag}^+(p)$. The Lie groups $SO(p)$ and $\mathrm{Diag}^+(p)$ carry natural bi-invariant Riemannian metrics g_{SO} and $g_{\mathcal{D}^+}$, defined as follows. The tangent space at U of $SO(p)$ is $T_U(SO(p)) = \{AU : A \in \mathfrak{so}(p)\}$, where $\mathfrak{so}(p)$ is the space of $p \times p$ antisymmetric matrices. At an arbitrary point $U \in SO(p)$ we define $g_{SO}\,|_U\,(A_1, A_2) = -\frac{1}{2}\mathrm{tr}(A_1 U^T A_2 U^T)$ for $A_1, A_2 \in \mathfrak{so}(p)$, where $\mathrm{tr}(A)$ is the trace of the matrix A. The tangent space $T_D(\mathrm{Diag}^+(p)) = \{LD : L \in \mathrm{Diag}(p)\}$ is canonically identified with $\mathrm{Diag}(p)$, the set of $p \times p$ diagonal matrices, and we define $g_{\mathcal{D}^+}|_D\,(L_1,L_2) = \mathrm{tr}(L_1D^{-1}L_2D^{-1})$, for $L_1,L_2 \in \mathrm{Diag}(p)$. Given eigen-decompositions (U_1,D_1) and (U_2,D_2) of SPD matrices X_1 and X_2, we measure the distance between their eigen-decompositions using the following product metric:

Definition 2.1. We define the *distance function* d_M on $M(p)$, with a weighting parameter $k > 0$, by

$$d_M^2((U_1,D_1),(U_2,D_2)) = k d_{SO}^2(U_1,U_2) + d_{\mathcal{D}^+}^2(D_1,D_2), \tag{2.1}$$

where $d_{SO}(U_1,U_2) = \frac{1}{\sqrt{2}}\|\mathrm{Log}(U_2U_1^{-1})\|_F$, $d_{\mathcal{D}^+}(D_1,D_2) = \|\mathrm{Log}(D_2D_1^{-1})\|_F$, and $\|.\|_F$ denotes the Frobenius norm.

In Definition 2.1 and (2.2) below, $\mathrm{Exp}(A)$ stands for the matrix exponential of A, and $\mathrm{Log}(R)$ for the principal matrix logarithm of R.[1] The weighting parameter k is a fixed constant throughout, and allows for combining two intrinsically incommensurate quantities with different relative weights (*cf.* Jung, Schwartzman and Groisser, 2015). Our theoretical results hold for any $k > 0$, but the choice of k affects certain results concerning, for instance, the diameter of the support of data. The parameter is set to $k = 1$ in all numerical results in this article, but can be adjusted in practice. As an example, if one wishes to standardize the variability in eigenvectors with respect to the eigenvalues, one may set k to be the ratio of the (suitably defined) Fréchet variance of eigenvalue matrices to the Fréchet variance of eigenvector matrices. The function d_M is the geodesic distance function determined by the Riemannian metric g_M that we define below.

For a geometric interpretation of the geodesic distance, note that the geodesic distance between eigen-decompositions (U_1,D_1) and (U_2,D_2) equals the length of the $M(p)$-valued geodesic

$$\gamma_{(U_1,D_1),(U_2,D_2)}(t) = (\mathrm{Exp}(t\mathrm{Log}(U_2U_1^{-1}))U_1, \mathrm{Exp}(t\mathrm{Log}(D_2D_1^{-1}))D_1) \tag{2.2}$$

connecting (U_1,D_1) and (U_2,D_2), which is a minimal-length smooth curve connecting these two points when the tangent spaces of $M(p)$ are equipped with the canonical inner product $g_M = kg_{SO} \oplus g_{\mathcal{D}^+}$. The functions $d_{\mathcal{D}^+}(D_1,D_2)$ and $d_{SO}(U_1,U_2)$ in (2.1) have the following interpretations: $d_{\mathcal{D}^+}(D_1,D_2)$ computes the Euclidean distance between $\mathrm{Log}(D_1)$ and $\mathrm{Log}(D_2)$, while $d_{SO}(U_1,U_2)$ equals the magnitude of the rotation angle of $U_2U_1^{-1}$ when $p = 2,3$.

The exponential map at $(U,D) \in M(p)$ is $\mathrm{Exp}_{(U,D)} : T_{(U,D)}M(p) \to M(p)$, given by

$$\mathrm{Exp}_{(U,D)}((AU,LD)) = (\mathrm{Exp}(A)U, \mathrm{Exp}(L)D). \tag{2.3}$$

The inverse of the exponential map at $(U,D) \in M(p)$, defined for $\mathcal{U}_{(U,D)} = \{(V,\Lambda) \in M(p) : \|\mathrm{Log}(VU^T)\|_F < \pi\}$, is $\mathrm{Log}_{(U,D)} : \mathcal{U}_{(U,D)} \to T_{(U,D)}M(p)$, and is given by

$$\mathrm{Log}_{(U,D)}((V,\Lambda)) = (\mathrm{Log}(VU^T)U, \mathrm{Log}(\Lambda D^{-1})D). \tag{2.4}$$

With the Riemannian metric $g_M = kg_{SO} \oplus g_{\mathcal{D}^+}$, the induced norm on the tangent space $T_{(U,D)}M(p)$ satisfies

$$\|(A_1U,L_1D) - (A_2U,L_2D)\|_{(U,D)}^2 = \frac{k}{2}\mathrm{tr}((A_1 - A_2)(A_1 - A_2)^T) + \mathrm{tr}((L_1 - L_2)(L_1 - L_2)^T),$$

for any two tangent vectors $(A_1U,L_1D),(A_2U,L_2D) \in T_{(U,D)}M(p)$.

[1] The principal logarithm for rotation matrices is defined on the set $\{R \in SO(p) : R \text{ is not an involution}\}$, a dense open subset of $SO(p)$. When there exists no principal logarithm of R, the notation $\mathrm{Log}(R)$ denotes any solution $A \in \mathfrak{so}(p)$ of $\mathrm{Exp}(A) = R$ satisfying that $\|A\|_F$ is the smallest among all such choices of A. For such rare cases, the geodesic (2.2) is not unique, but $\|\mathrm{Log}(R)\|_F$ is well defined.

2.2. Minimal smooth scaling-rotation curves and scaling-rotation distance

Since eigen-decompositions are not unique, any method for computing the distance between SPD matrices using the eigen-decomposition space must take this non-uniqueness into account. To address this, Jung, Schwartzman and Groisser (2015) proposed the following distance for $\mathrm{Sym}^+(p)$:

Definition 2.2. Let $\mathcal{F} : M(p) \to \mathrm{Sym}^+(p)$ denote the eigen-composition map $\mathcal{F}(U,D) = UDU^T$, and for any $X \in \mathrm{Sym}^+(p)$, let $\mathcal{F}^{-1}(X)$ denote the set of eigen-decompositions of X. The *scaling-rotation distance* between $X \in \mathrm{Sym}^+(p)$ and $Y \in \mathrm{Sym}^+(p)$ is

$$d_{\mathcal{SR}}(X,Y) = \inf_{\substack{(U_X,D_X)\in\mathcal{F}^{-1}(X),\\ (U_Y,D_Y)\in\mathcal{F}^{-1}(Y)}} d_M((U_X,D_X),(U_Y,D_Y)).$$

Eigen-decompositions $(U_X^*, D_X^*) \in \mathcal{F}^{-1}(X)$ and $(U_Y^*, D_Y^*) \in \mathcal{F}^{-1}(Y)$ form a *minimal pair* if

$$d_M((U_X^*, D_X^*),(U_Y^*, D_Y^*)) = d_{\mathcal{SR}}(X,Y).$$

Remark 2.3. Since the sets $\mathcal{F}^{-1}(X)$ and $\mathcal{F}^{-1}(Y)$ are compact for any $X,Y \in \mathrm{Sym}^+(p)$, there will always be a pair of eigen-decompositions of X and Y that form a minimal pair.

Remark 2.4. The function $d_{\mathcal{SR}}$ is not a true metric on $\mathrm{Sym}^+(p)$ since there are instances in which the triangle inequality fails. It is a semi-metric and invariant under simultaneous matrix inversion, uniform scaling and conjugation by a rotation matrix (Jung, Schwartzman and Groisser, 2015, Theorem 3.11). When restricted to the subset of SPD matrices which possess no repeated eigenvalues, $d_{\mathcal{SR}}$ is a true metric (Jung, Schwartzman and Groisser, 2015, Theorem 3.12).

For SPD matrices X,Y and their eigen-decompositions $(U_X,D_X) \in \mathcal{F}^{-1}(X)$, $(U_Y,D_Y) \in \mathcal{F}^{-1}(Y)$, one can create a smooth scaling-rotation (SSR) curve on $\mathrm{Sym}^+(p)$ connecting X and Y as $\chi_{X,Y}(t) = \mathcal{F}(\gamma_{(U_X,D_X),(U_Y,D_Y)}(t))$, where $\gamma_{(U_X,D_X),(U_Y,D_Y)}(t)$ is a minimal-length geodesic curve defined in (2.2). If one considers the family of all possible geodesics in $(M(p), g_M)$ from $\mathcal{F}^{-1}(X)$ to $\mathcal{F}^{-1}(Y)$, the scaling-rotation distance equals the length of the shortest geodesics in that family. By definition, the shortest geodesic (which may not be uniquely defined) connects a minimal pair $(U_X^*, D_X^*) \in \mathcal{F}^{-1}(X)$ and $(U_Y^*, D_Y^*) \in \mathcal{F}^{-1}(Y)$. The weighting parameter k, introduced in Definition 2.1, affects the relative lengths of these geodesics. Thus, which geodesic becomes the shortest geodesic depends on the value of k. (See supplementary Section 2 of Jung, Schwartzman and Groisser (2015)). Computing $d_{\mathcal{SR}}(X,Y)$ for any dimension p is straightforward when X and Y both have no repeated eigenvalues, since X and Y then both have finitely many eigen-decompositions and therefore finitely many connecting SSR curves, or when one of X,Y is a scaled identity matrix. Formulas for computing $d_{\mathcal{SR}}(X,Y)$ for all possible eigenvalue-multiplicity combinations of arguments X and Y are provided in Groisser, Jung and Schwartzman (2017) for $p = 2,3$.

2.3. The stratification of $\mathrm{Sym}^+(p)$ and fibers of the eigen-composition map

The space $\mathrm{Sym}^+(p)$ is naturally stratified by eigenvalue-multiplicity types, identified naturally with partitions of p. (A *partition* of p, generally denoted symbolically in the form "$k_1 + k_2 + \cdots + k_r$", is a finite non-decreasing sequence of positive integers, called *parts* of the partition, whose sum is p. The parts k_i of such a partition label eigenvalue-multiplicities; the stratum corresponding to the partition

$k_1 + k_2 + \cdots + k_r$ has r distinct eigenvalues, whose greatest multiplicity is k_1, next-greatest multiplicity is k_2, etc. Only the *multiplicities* of eigenvalues, not their relative sizes, are relevant to this labeling.) We will use the notation S_p^{top} to denote the subset of SPD matrices which have no repeated eigenvalues (the superscript "top" refers to the "top stratum"); this corresponds to the partition $1 + 1 + \cdots + 1$. We also use the notation $S_p^{\text{lwr}} := \text{Sym}^+(p) \setminus S_p^{\text{top}}$, for the union of all "lower" strata, and $S_p^{\text{bot}} \subset S_p^{\text{lwr}}$ denotes the set of SPD matrices having just a single, multiplicity-p eigenvalue. If $X, Y \in \text{Sym}^+(p)$ are in the same stratum, then $\mathcal{F}^{-1}(X)$ and $\mathcal{F}^{-1}(Y)$ are diffeomorphic, as we elaborate below.

Let $\mathcal{S}_p$ be the group of permutations of $\{1, 2, \ldots, p\}$. For a permutation $\pi \in \mathcal{S}_p$ and $D \in \text{Diag}^+(p)$, the natural left action of $\mathcal{S}_p$ on $\text{Diag}^+(p)$ is given by permuting the diagonal entries of D. We call a $p \times p$ matrix P a *signed-permutation matrix* if for some $\pi \in \mathcal{S}_p$ the entries of P satisfy $P_{ij} = \pm\delta_{i,\pi(j)}$. We call such P *even* if $\det(P) = 1$. The set $\mathcal{G}(p)$ of all such *even signed-permutation matrices* has exactly $2^{p-1}p!$ elements, and is a matrix subgroup of $SO(p)$. The natural left-action of $\mathcal{G}(p)$ on $M(p)$ is given by

$$h \cdot (U, D) := (Uh^{-1}, h \cdot D), \tag{2.5}$$

where $h \in \mathcal{G}(p)$ and $h \cdot D := hDh^{-1}$. The action of h on (U, D) represents the simultaneous permutation (by the unsigned permutation associated with h) of columns of U and diagonal elements of D, and the sign-changes of the columns of U. The identity element of $\mathcal{G}(p)$ is I_p.

It is shown in Jung, Schwartzman and Groisser (2015) that the fiber $\mathcal{F}^{-1}(X)$—that is, the set of eigen-decompositions of X—can be expressed as

$$\mathcal{F}^{-1}(X) = \{h \cdot (UR, D) : R \in G_D, h \in \mathcal{G}(p)\}, \tag{2.6}$$

with any $(U, D) \in \mathcal{F}^{-1}(X)$, where $G_D = \{R \in SO(p) : RDR^T = D\}$. Thus, the action of $\mathcal{G}(p)$ on $M(p)$ is fiber-preserving.

The topological structure of the fiber $\mathcal{F}^{-1}(X)$ depends on the stratum to which X belongs. If $X \in S_p^{\text{top}}$, then for any eigen-decomposition $(U, D) \in \mathcal{F}^{-1}(X)$, the orbit

$$\mathcal{G}(p) \cdot (U, D) = \{h \cdot (U, D) : h \in \mathcal{G}(p)\} \tag{2.7}$$

is *exactly* $\mathcal{F}^{-1}(X)$, the set of eigen-decompositions of X. Intuitively, any eigen-decomposition of $X \in S_p^{\text{top}}$ can be obtained from any other by a sign-change of eigenvectors and a simultaneous permutation of eigenvectors and eigenvalues. In contrast, if $X \in S_p^{\text{bot}}$ (i.e., X is a scaled identity matrix), then $G_X = SO(p)$ and $h \cdot X = X$ for all $h \in \mathcal{G}(p)$, thus the fiber of $\mathcal{F}$ at X is $\mathcal{F}^{-1}(X) = SO(p) \times \{X\}$. Fibers other than these two extremes are more complicated to describe, but a complete characterization of all the fibers of $\mathcal{F}$ can be found in Groisser, Jung and Schwartzman (2017).

3. Location estimation under the scaling-rotation framework

3.1. Fréchet mean

An approach often used for developing location estimators for non-Euclidean metric spaces is Fréchet mean estimation (Fréchet, 1948), in which estimators are derived as minimizers of a metric-dependent sample mean-squared error.

Definition 3.1. Let M be a metric space with metric ρ and suppose that $X, X_1, \ldots, X_n$ are i.i.d. M-valued random variables with induced probability measure P on M. The *population Fréchet mean set* is

$$\operatorname*{argmin}_{C \in M} \int_M \rho^2(X, C) P(dX).$$

The *sample Fréchet mean set* is

$$\operatorname*{argmin}_{C \in M} \frac{1}{n} \sum_{i=1}^{n} \rho^2(X_i, C).$$

Examples of location estimators that have been developed for $\mathrm{Sym}^+(p)$ using the sample Fréchet mean estimation framework include the log-Euclidean mean (Arsigny et al., 2006/07), affine-invariant mean (Fletcher et al., 2004, Pennec, Fillard and Ayache, 2006), Procrustes size-and-shape mean (Dryden, Koloydenko and Zhou, 2009), and the log-Cholesky average (Lin, 2019). Below, we allow ourselves to use the "Fréchet mean" terminology of Definition 3.1 when the metric space (M, ρ) is replaced by the semi-metric space $(\mathrm{Sym}^+(p), d_{\mathcal{SR}})$.

3.2. Scaling-rotation means

We now define the population and sample scaling-rotation mean sets, consisting of the Fréchet means of SPD matrices under the scaling-rotation framework. Let P be a Borel probability measure on $\mathrm{Sym}^+(p)$, and $X_1, \ldots, X_n$ be deterministic data points in $\mathrm{Sym}^+(p)$. Note that Borel measures on $\mathrm{Sym}^+(p)$ include both discrete and absolutely continuous measures, as well as mixtures of those.

Definition 3.2. The *population scaling-rotation (SR) mean set* with respect to P is

$$E^{(\mathcal{SR})} := \operatorname*{argmin}_{S \in \mathrm{Sym}^+(p)} f^{(\mathcal{SR})}(S), \quad f^{(\mathcal{SR})}(S) = \int_{\mathrm{Sym}^+(p)} d^2_{\mathcal{SR}}(X, S) P(dX). \tag{3.1}$$

Given $X_1, \ldots, X_n \in \mathrm{Sym}^+(p)$, the *sample SR mean set* is

$$E_n^{(\mathcal{SR})} := \operatorname*{argmin}_{S \in \mathrm{Sym}^+(p)} f_n^{(\mathcal{SR})}(S), \quad f_n^{(\mathcal{SR})}(S) = \frac{1}{n} \sum_{i=1}^{n} d^2_{\mathcal{SR}}(X_i, S).$$

Since, for some $S \in \mathrm{Sym}^+(p)$, the function $d_{\mathcal{SR}}(\cdot, S) : \mathrm{Sym}^+(p) \to \mathbb{R}$ has discontinuities (see Section A of the Supplementary Material (Jung et al., 2024)), we must address whether the objective function $f^{(\mathcal{SR})}$ of (3.1) is well-defined. We defer this discussion to Section 4.1.

Locating a sample SR mean can be recast as solving a difficult constrained optimization problem on $M(p)^n$ since

$$\frac{1}{n} \sum_{i=1}^{n} d^2_{\mathcal{SR}}(X_i, S) = \frac{1}{n} \sum_{i=1}^{n} d^2_{M}((U_i^*, D_i^*), (U_S^{*,i}, D_S^{*,i})), \tag{3.2}$$

where for each $i = 1, \ldots, n$, $(U_i^*, D_i^*) \in \mathcal{F}^{-1}(X_i)$ and $(U_S^{*,i}, D_S^{*,i}) \in \mathcal{F}^{-1}(S)$ are an arbitrary minimal pair. Due to the non-uniqueness of eigen-decompositions, there may be many pairs of eigen-decompositions of X_i and S which form a minimal pair.

However, when $S \in S_p^{\text{top}}$ the scaling-rotation distance simplifies to

$$d_{\mathcal{SR}}(X,S) = \inf_{(U_X,D_X)\in\mathcal{F}^{-1}(X)} d_M((U_X,D_X),(U_S,D_S)), \tag{3.3}$$

where (U_S,D_S) is *any* eigen-decomposition of S. In this case, $d_{\mathcal{SR}}(X,S)$ is easier to compute since one can select an arbitrary eigen-decomposition (U_S,D_S) of S and then determine the infimum of the distances between (U_S,D_S) and the eigen-decompositions of X. If S has repeated eigenvalues (or, equivalently, S is in a lower stratum), this simplification does not hold in general; there may be no eigen-decomposition of S that is at minimal distance from $\mathcal{F}^{-1}(X_i)$ simultaneously for all i.

From the simplification in (3.3), we propose to solve for minimizers of the simplified objective function

$$(U,D) \mapsto \frac{1}{n}\sum_{i=1}^{n} \inf_{(U_X,D_X)\in\mathcal{F}^{-1}(X_i)} d_M^2((U_X,D_X),(U,D)),$$

where the argument (U,D) is an arbitrarily chosen eigen-decomposition of the argument S from (3.2).

To formally define this simplified optimization problem, we first define the following measure of distance between an SPD matrix and a given eigen-decomposition of another SPD matrix:

Definition 3.3. The *partial scaling-rotation (PSR) distance* is the map $d_{\mathcal{PSR}} : \text{Sym}^+(p) \times M(p) \to [0,\infty)$ given by

$$d_{\mathcal{PSR}}(X,(U,D)) = \inf_{(U_X,D_X)\in\mathcal{F}^{-1}(X)} d_M((U_X,D_X),(U,D)).$$

It can be checked from the definitions that for any $X \in \text{Sym}^+(p)$ and any $(U,D) \in M(p)$

$$d_{\mathcal{SR}}(X,\mathcal{F}(U,D)) \le d_{\mathcal{PSR}}(X,(U,D)), \tag{3.4}$$

and by (3.3), the equality in (3.4) holds if $\mathcal{F}(U,D) \in S_p^{\text{top}}$.

Definition 3.4. The population and sample *partial scaling-rotation (PSR) mean sets* are subsets of $M(p)$ and are defined respectively by $E^{(\mathcal{PSR})} := \text{argmin}_{(U,D)\in M(p)} f^{(\mathcal{PSR})}(U,D)$ and $E_n^{(\mathcal{PSR})} := \text{argmin}_{(U,D)\in M(p)} f_n^{(\mathcal{PSR})}(U,D)$, where

$$f^{(\mathcal{PSR})}(U,D) = \int_{\text{Sym}^+(p)} d_{\mathcal{PSR}}^2(X,(U,D))P(dX), \tag{3.5}$$

$$f_n^{(\mathcal{PSR})}(U,D) = \frac{1}{n}\sum_{i=1}^{n} d_{\mathcal{PSR}}^2(X_i,(U,D)).$$

In Sections 4.1 and 4.2 we show that for any Borel probability measure on $\text{Sym}^+(p)$, the population mean set $E^{(\mathcal{PSR})}$ is well-defined and non-empty. There, we also show that both $E_n^{(\mathcal{PSR})}$ and $E_n^{(\mathcal{SR})}$ are non-empty for any sample $X_1,\dots,X_n$. An iterative algorithm to compute a sample PSR mean is given in Section 3.3.

The PSR means lie in $M(p)$ and can be mapped to $\text{Sym}^+(p)$ via the eigen-composition map. The PSR mean set can be thought of as yielding an approximation of the SR mean set, and it is of interest to know when the two sets are "equivalent". The theorem below provides conditions under which $E^{(\mathcal{PSR})} \subset M(p)$ is equivalent to $E^{(\mathcal{SR})} \subset \text{Sym}^+(p)$ in the sense that every member of $E^{(\mathcal{PSR})}$ is an eigen-decomposition of a member of $E^{(\mathcal{SR})}$ and vice-versa.

Theorem 3.5. *Suppose that $f^{(\mathcal{PSR})}(U,D) < \infty$ for some $(U,D) \in M(p)$. If $E^{(\mathcal{SR})} \subset S_p^{\text{top}}$, then*

$$E^{(\mathcal{PSR})} = \mathcal{F}^{-1}(E^{(\mathcal{SR})}) \text{ and } \mathcal{F}(E^{(\mathcal{PSR})}) = E^{(\mathcal{SR})}.$$

The statement above holds when $E^{(\mathcal{SR})}$ and $E^{(\mathcal{PSR})}$ are replaced by $E_n^{(\mathcal{SR})}$ and $E_n^{(\mathcal{PSR})}$, respectively.

The previous theorem suggests that in many realistic situations, there may be no cost to using the PSR means in place of the SR means, which are more difficult to compute in practice. These realistic situations include the case where all SR means have distinct eigenvalues. If minimizing $f^{(\mathcal{SR})}$ or $f_n^{(\mathcal{SR})}$ over S_p^{lwr} (the union of lower strata of $\text{Sym}^+(p)$) is feasible, then Theorem 3.6 below can be used to tell whether a PSR mean is equivalent to an SR mean.

For the rest of the paper, we generally use the notation m rather than (U,D) for an arbitrary element of $M(p)$ if there is no explicit need for writing out the eigenvector and eigenvalue matrices separately.

Theorem 3.6. *Let $m^{\mathcal{PSR}} \in M(p)$ be a PSR mean with respect to a probability measure P on $\text{Sym}^+(p)$.*

(a) *If $f^{(\mathcal{SR})}(\mathcal{F}(m^{\mathcal{PSR}})) \le \min_{S \in S_p^{\text{lwr}}} f^{(\mathcal{SR})}(S)$, then $\mathcal{F}(m^{\mathcal{PSR}}) \in E^{(\mathcal{SR})}$.*

(b) *If $f^{(\mathcal{SR})}(\mathcal{F}(m^{\mathcal{PSR}})) > \min_{S \in S_p^{\text{lwr}}} f^{(\mathcal{SR})}(S)$, then $\mathcal{F}(m^{\mathcal{PSR}}) \notin E^{(\mathcal{SR})}$ and $E^{(\mathcal{SR})} \subset S_p^{\text{lwr}}$.*

Let $\hat{m}^{\mathcal{PSR}} \in M(p)$ be a sample PSR mean with respect to a given sample $X_1, \ldots, X_n \in \text{Sym}^+(p)$. Similarly to the statements above,

(c) *If $f_n^{(\mathcal{SR})}(\mathcal{F}(\hat{m}^{\mathcal{PSR}})) \le \min_{S \in S_p^{\text{lwr}}} f_n^{(\mathcal{SR})}(S)$, then $\mathcal{F}(\hat{m}^{\mathcal{PSR}}) \in E_n^{(\mathcal{SR})}$.*

(d) *If $f_n^{(\mathcal{SR})}(\mathcal{F}(\hat{m}^{\mathcal{PSR}})) > \min_{S \in S_p^{\text{lwr}}} f_n^{(\mathcal{SR})}(S)$, then $\mathcal{F}(\hat{m}^{\mathcal{PSR}}) \notin E_n^{(\mathcal{SR})}$ and $E_n^{(\mathcal{SR})} \subset S_p^{\text{lwr}}$.*

We remark that for $p = 2$, $S_p^{\text{lwr}} = \{cI_2 : c > 0\}$ and the function $f_n^{(\mathcal{SR})}$ can be efficiently minimized over S_p^{lwr} by a one-dimensional numerical optimization.

A key condition that enables us to ensure the equivalence of the PSR means to SR means is that all SR means have no repeated eigenvalues (i.e., $E^{(\mathcal{SR})} \subset S_p^{\text{top}}$). This condition, of course, depends on the distribution P. Below, we give a condition on P sufficient to ensure that $E^{(\mathcal{SR})} \subset S_p^{\text{top}}$. To express this condition, let $\delta : \text{Sym}^+(p) \to [0,\infty)$ be $\delta(S) = \inf\{d_{\mathcal{SR}}(S,S') : S' \in S_p^{\text{lwr}}\}$. Thus, $\delta(S)$ is a “distance” from S to lower strata of $\text{Sym}^+(p)$. (Because S_p^{lwr} is closed, $\delta(S) > 0$ for any $S \in S_p^{\text{top}}$.)

Theorem 3.7. *Let X be a $\text{Sym}^+(p)$-valued random variable with distribution P. Assume that there exists $S_0 \in S_p^{\text{top}}$ and $r \in (0, \delta(S_0)/3)$ such that*

$$P(d_{\mathcal{SR}}(X,S_0) \le r) = 1.$$

Then $E^{(\mathcal{SR})} \subset S_p^{\text{top}}$. As a corollary, if $X_1, \ldots, X_n \in \text{Sym}^+(p)$ and there exists $S_0 \in S_p^{\text{top}}$ and $r \in (0, \delta(S_0)/3)$ satisfying $d_{\mathcal{SR}}(X_i,S_0) \le r$ for $i = 1, \ldots, n$, then $E_n^{(\mathcal{SR})} \subset S_p^{\text{top}}$.

The condition of Theorem 3.7 requires that the support of P be sufficiently far from lower strata of $\text{Sym}^+(p)$. Since lower strata consist of SPD matrices with two or more equal eigenvalues, this condition requires that all SPD matrices in the support have sufficiently large gaps between eigenvalues. But this condition is not necessary for the conclusion. The condition, however, cannot be weakened to the condition that all data lie in S_p^{top} (for finite samples); there are examples in which this weaker condition is met, but $E_n^{(\mathcal{SR})} \subset S_p^{\text{lwr}}$ (see numerical examples in Section C.1 of the Supplementary Material (Jung et al., 2024)).

3.3. Sample PSR mean estimation algorithm

Given a sample $X_1, \ldots, X_n \in \mathrm{Sym}^+(p)$, we propose an algorithm for approximating a member of $E_n^{(\mathcal{PSR})}$, that is to find a minimizer of $f_n^{(\mathcal{PSR})}$. The algorithm is similar to the generalized Procrustes algorithm (Gower, 1975).

Procedure 3.8 (Sample PSR Mean). Set tolerance $\varepsilon > 0$ and pick initial guess $(\hat{U}^{(0)}, \hat{D}^{(0)}) \in M(p)$. Set $j = 0$.

Step 1. For $i = 1, \ldots, n$, find $(U_i^{(j)}, D_i^{(j)}) \in \mathcal{F}^{-1}(X_i)$ that has the smallest geodesic distance from $(\hat{U}^{(j)}, \hat{D}^{(j)})$.

Step 2. Compute $(\hat{U}^{(j+1)}, \hat{D}^{(j+1)}) \in \mathrm{argmin}_{(U,D)\in M(p)} \frac{1}{n}\sum_{i=1}^n d_M^2((U_i^{(j)}, D_i^{(j)}), (U, D))$.

If $|f_n^{(\mathcal{PSR})}(\hat{U}^{(j+1)}, \hat{D}^{(j+1)}) - f_n^{(\mathcal{PSR})}(\hat{U}^{(j)}, \hat{D}^{(j)})| > \varepsilon$, increment j and repeat Steps 1 and 2. Otherwise, $(\hat{U}_{\mathcal{PSR}}, \hat{D}_{\mathcal{PSR}}) = (\hat{U}^{(j+1)}, \hat{D}^{(j+1)})$ is the approximate sample PSR mean produced by this algorithm, given the tolerance ε and initial guess $(\hat{U}^{(0)}, \hat{D}^{(0)})$.

Remark 3.9. The above procedure will always terminate since each step of the procedure reduces the value of the objective function, i.e., $f_n^{(\mathcal{PSR})}(\hat{U}^{(j+1)}, \hat{D}^{(j+1)}) \le f_n^{(\mathcal{PSR})}(\hat{U}^{(j)}, \hat{D}^{(j)})$ for any $j \ge 0$, and $f_n^{(\mathcal{PSR})}(U, D)$ is bounded below by zero for any $(U, D) \in M(p)$.

If X_i lies in S_p^{top}, performing Step 1 will simply require searching over the $2^{(p-1)}p!$ distinct eigen-decompositions of X_i to find one that attains the minimal geodesic distance from $(\hat{U}^{(j)}, \hat{D}^{(j)})$. Solving for the minimizing eigen-decomposition of X_i is also easy if X_i is a scaled identity matrix ($X_i \in S_p^{\mathrm{bot}}$), since the fact that $X_i = cI_p = U(cI_p)U^T$ for any $U \in SO(p)$ implies that $(\hat{U}^{(j)}, cI_p)$ will be the eigen-decomposition of X_i with minimal geodesic distance from $(\hat{U}^{(j)}, \hat{D}^{(j)})$. Determining the minimizing eigen-decomposition of X_i when $p = 3$ and X_i has two distinct eigenvalues can be done by comparing three closed-form expressions, as described in Groisser, Jung and Schwartzman (2017). For $p > 3$, there are no known corresponding closed-form expressions for determining a minimizing eigen-decomposition of $X_i \in S_p^{\mathrm{lwr}} \setminus S_p^{\mathrm{bot}}$.

The optimization problem over $M(p)$ in Step 2 can be divided into separate minimization problems over $\mathrm{Diag}^+(p)$ and $SO(p)$:

$$\hat{D}^{(j+1)} = \underset{D\in \mathrm{Diag}^+(p)}{\mathrm{argmin}} \frac{1}{n}\sum_{i=1}^{n} \|\mathrm{Log}(D_i^{(j)}) - \mathrm{Log}(D)\|_F^2,$$

$$\hat{U}^{(j+1)} \in \underset{U\in SO(p)}{\mathrm{argmin}} \frac{1}{n}\sum_{i=1}^{n} \|\mathrm{Log}(U_i^{(j)}U^{-1})\|_F^2.$$

The solution $\hat{D}^{(j+1)}$ is uniquely given by $\hat{D}^{(j+1)} = \mathrm{Exp}\{\frac{1}{n}\sum_{i=1}^n \mathrm{Log}(D_i^{(j)})\}$, while $\hat{U}^{(j+1)}$ usually must be approximated via numerical procedures. It is shown in Manton (2004) that when the rotation matrices $U_1^{(j)}, \ldots, U_n^{(j)}$ lie within a geodesic ball of radius $\frac{\pi}{2}$, there is a unique minimizer $\hat{U}^{(j+1)}$, and this minimizer can be approximated by a globally convergent gradient descent algorithm on $(SO(p), g_{SO})$. It is highly unlikely that one would be able to de-couple estimation of the eigenvalue and eigenvector means in this manner while solving for a sample SR mean.

4. Theoretical properties of scaling-rotation means

4.1. Lower semicontinuity and other properties of $d_{\mathcal{SR}}$ and $d_{\mathcal{PSR}}$

One of the complications in using the SR framework is that the symmetric function $d_{\mathcal{SR}}$ is not continuous in either variable (see Section A of the Supplementary Material (Jung et al., 2024) for an example). Unfortunately, $d_{\mathcal{PSR}}$ is also not continuous at every point of $\mathrm{Sym}^+(p) \times M(p)$, as illustrated by the following example. Let $X(\varepsilon) := \mathrm{diag}(e^{\varepsilon}, e^{-\varepsilon})$ and $(U,D) = (R(\theta), I_2)$, where $R(\theta)$ is the 2×2 rotation matrix corresponding to a counterclockwise rotation by angle θ. Then for any $\varepsilon \neq 0$ and $0 < |\theta| < \pi/4$,

$$d_{\mathcal{PSR}}(X(\varepsilon),(U,D)) = (k\theta^2 + 2\varepsilon^2)^{1/2},$$

which implies that $d_{\mathcal{PSR}}(X(\varepsilon),(U,D)) \to \sqrt{k}|\theta|$ as $\varepsilon \to 0$. Since $d_{\mathcal{PSR}}(X(0),(U,D)) = 0$, it follows that $d_{\mathcal{PSR}}$ is not continuous at $(I_2,(U,D))$, and therefore $d_{\mathcal{PSR}}$ is not continuous on $\mathrm{Sym}^+(p) \times M(p)$. Nevertheless, $d_{\mathcal{PSR}}$ is continuous with respect to the second variable in $M(p)$, and is jointly continuous on $S_p^{\mathrm{top}} \times M(p)$, as we state below.

Lemma 4.1.

(a) *$d_{\mathcal{PSR}}$ is continuous on $S_p^{\mathrm{top}} \times M(p)$.*

(b) *For each $S \in \mathrm{Sym}^+(p)$, the function $d_{\mathcal{PSR}}(S,\cdot) : M(p) \to [0,\infty)$ is Lipschitz, with Lipschitz-constant 1. That is, for all $m_1, m_2 \in M(p)$,*

$$|d_{\mathcal{PSR}}(S,m_1) - d_{\mathcal{PSR}}(S,m_2)| < d_M(m_1,m_2).$$

In particular, $d_{\mathcal{PSR}}(S,\cdot)$ is uniformly continuous for each S.

Since both $d_{\mathcal{SR}}$ and $d_{\mathcal{PSR}}$ are not continuous, in principle we do not know yet whether the integrals of $d^2_{\mathcal{SR}}(\cdot,\Sigma)$ and $d^2_{\mathcal{PSR}}(\cdot,(U,D))$, for $\Sigma \in \mathrm{Sym}^+(p)$ and $(U,D) \in M(p)$, in Definitions 3.2 and 3.4, are well defined. A related question is: under which conditions do the population (partial) scaling-rotation means exist? A key observation in answering these questions is that these functions $d^2_{\mathcal{SR}}(\cdot,\Sigma)$ and $d^2_{\mathcal{PSR}}(\cdot;(U,D))$ are *lower semicontinuous* (LSC). (Recall that a function $f : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X}$ is a topological space, is LSC at a point $x_0 \in \mathcal{X}$ if for all $\epsilon > 0$, there exists an open neighborhood $\mathcal{U}$ of x_0 such that $f(x) > f(x_0) - \epsilon$ for all $x \in \mathcal{U}$. If f is LSC at each $x_0 \in \mathcal{X}$, we say that f is LSC.)

Definition 4.2. Let $\mathcal{X}$ be a topological space and $\mathcal{Y}$ be a set, and let $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.

(a) We say that f is *LSC in its first variable, uniformly with respect to its second variable*, if for all $x_0 \in \mathcal{X}$ and $\epsilon > 0$, there exists an open neighborhood $\mathcal{U}$ of x_0 such that

$$f(x,y) > f(x_0,y) - \epsilon \quad \text{for all } x \in \mathcal{U} \text{ and all } y \in \mathcal{Y}. \tag{4.1}$$

(b) If $\mathcal{Y}$ is also a topological space, we say that f *is LSC in its first variable, locally uniformly with respect to its second variable*, if every $y_0 \in \mathcal{Y}$ has an open neighborhood $\mathcal{V}$ such that $f|_{\mathcal{X}\times\mathcal{V}}$ is LSC in the first variable, uniformly with respect to the second. If $\mathcal{Y}$ is locally compact, this property is equivalent to: for every compact set $K \subset \mathcal{Y}$, $f|_{\mathcal{X}\times K}$ is LSC in the first variable, uniformly with respect to the second.

Any finite-dimensional manifold (in particular, $M(p)$) is locally compact.

Theorem 4.3.

(a) *Let* $S_0 \in \mathrm{Sym}^+(p)$*, and* $m_0 \in M(p)$*. Then the functions* $d^2_{\mathcal{SR}}(\cdot, S_0)$ *and* $d^2_{\mathcal{PSR}}(\cdot, m_0)$ *and their square-roots are LSC.*

(b) *The functions* $d_{\mathcal{SR}}(\cdot,\cdot)$*,* $d^2_{\mathcal{SR}}(\cdot,\cdot)$*,* $d_{\mathcal{PSR}}(\cdot,\cdot)$ *and* $d^2_{\mathcal{PSR}}(\cdot,\cdot)$ *are LSC in the first variable, locally uniformly with respect to the second variable.*

In this theorem, part (a) is actually redundant; it is a special case of part (b), with the one-point set $\{S_0\}$ playing the role of the compact set in Definition 4.2(b). Also, for $d_{\mathcal{SR}}$ and $d^2_{\mathcal{SR}}$, the terms "first variable" and "second variable" in Theorem 4.3 can be interchanged, since $d_{\mathcal{SR}}$ is symmetric. Verifying Theorem 4.3 requires substantial background work regarding the geometry of the eigen-decomposition space $M(p)$ and the eigen-composition map $\mathcal{F}$. The following lemma is the key technical result used in proving Theorem 4.3. The radius-r open ball centered at $m_0 \in M(p)$ is $B_r^{d_M}(m_0) := \{m \in M(p) : d_M(m, m_0) < r\}$.

Lemma 4.4. *Let* $K \subset \mathrm{Sym}^+(p)$ *be a compact set. Let* $\epsilon > 0$ *and let* $S \in \mathrm{Sym}^+(p)$*. There exists* $\delta_1 = \delta_1(S, K, \epsilon) > 0$ *such that for all* $S_0 \in K$*, all* $m_0 \in \mathcal{F}^{-1}(S_0)$*, all* $m \in \mathcal{F}^{-1}(S)$*, and all* $S' \in \mathcal{F}\big(B_{\delta_1}^{d_M}(m)\big)$*,*

$$d_{\mathcal{SR}}(S', S_0)^2 > d_{\mathcal{SR}}(S, S_0)^2 - \epsilon \tag{4.2}$$

and

$$d_{\mathcal{PSR}}(S', m_0)^2 > d_{\mathcal{PSR}}(S, m_0)^2 - \epsilon. \tag{4.3}$$

Lemma 4.4 does not immediately imply that $d_{\mathcal{SR}}(\cdot, S_0)$ or $d_{\mathcal{PSR}}(\cdot, m_0)$ is LSC at S, because the set $\mathcal{F}(B_{\delta_1}^{d_M}(m))$ in the lemma is not always open in $\mathrm{Sym}^+(p)$ ($\mathcal{F}$ does not map arbitrary open sets to open sets). However, there exists an open ball centered at $\mathcal{F}(m)$ in $\mathrm{Sym}^+(p)$ with radius smaller than δ_1 that is contained in $\mathcal{F}(B_{\delta_1}^{d_M}(m))$ (Corollary B.14). The background and our proofs of these supporting results and Theorem 4.3 are provided in Section B.2 of the Supplementary Material (Jung et al., 2024).

Semicontinuous real-valued functions are (Borel) measurable, so an immediate consequence of Theorem 4.3(a) is that the integrals defining the objective functions $f^{(\mathcal{SR})}$ and $f^{(\mathcal{PSR})}$ for the population (partial) scaling-rotation means exist in $\mathbb{R} \cup \{\infty\}$. This establishes the following.

Proposition 4.5. *Let P be any Borel probability measure on* $\mathrm{Sym}^+(p)$*.*

(i) *For any* $S \in \mathrm{Sym}^+(p)$*, the integral* $\int_{\mathrm{Sym}^+(p)} d^2_{\mathcal{SR}}(\cdot, S) dP$ *is well-defined in* $[0, \infty]$.

(ii) *For any* $m \in M(p)$*, the integral* $\int_{\mathrm{Sym}^+(p)} d^2_{\mathcal{PSR}}(\cdot, m) dP$ *is well-defined in* $[0, \infty]$.

(Proof for Proposition 4.5 is omitted.)

A finite-variance condition for the random variable $X \in \mathrm{Sym}^+(p)$ with respect to the (partial) scaling-rotation distance (already needed to define $E^{(\mathcal{SR})}$ and $E^{(\mathcal{PSR})}$) is required to establish (semi-)continuity of $f^{(\mathcal{SR})}$ and $f^{(\mathcal{PSR})}$, non-emptiness of $E^{(\mathcal{SR})}$ and $E^{(\mathcal{PSR})}$ (discussed in Section 4.2), and relationships between these sets (in Section 3.2). For a probability measure P on $\mathrm{Sym}^+(p)$, we say P has *finite SR-variance* if $f^{(\mathcal{SR})}(S) < \infty$ for all $S \in \mathrm{Sym}^+(p)$. Likewise, P has *finite PSR-variance* if $f^{(\mathcal{PSR})}(U, D) < \infty$ for all $(U, D) \in M(p)$. The following result shows that such a condition needs to be assumed only at a single point, rather than at all points.

Lemma 4.6. *Let P be a Borel probability measure on* $\mathrm{Sym}^+(p)$ *and let* $f^{(\mathcal{PSR})}$ *and* $f^{(\mathcal{SR})}$ *be the corresponding objective functions defined in equations* (3.5) *and* (3.1).

(a) *If* $f^{(\mathcal{PSR})}(m) < \infty$ *for some* $m \in M(p)$*, then* $f^{(\mathcal{PSR})}(m) < \infty$ *for any* $m \in M(p)$*, and* $f^{(\mathcal{SR})}(S) < \infty$ *for any* $S \in \mathrm{Sym}^+(p)$*.*
(b) *If* $f^{(\mathcal{SR})}(S) < \infty$ *for some* $S \in \mathrm{Sym}^+(p)$*, then* $f^{(\mathcal{SR})}(S) < \infty$ *for any* $S \in \mathrm{Sym}^+(p)$*.*

By (3.4), any probability measure with finite PSR-variance always has finite SR variance.

We conclude this background section by answering a natural question: Are the SR and PSR mean functions $f^{(\mathcal{SR})}$ and $f^{(\mathcal{PSR})}$ (semi-)continuous?

Lemma 4.7. *Let* P *be a Borel probability measure on* $\mathrm{Sym}^+(p)$*.*

(i) *If* P *is supported in a compact set* $K \subset \mathrm{Sym}^+(p)$*, then* $f^{(\mathcal{SR})} : \mathrm{Sym}^+(p) \to \mathbb{R}$ *is LSC.*
(ii) *If* P *has finite PSR-variance, then* $f^{(\mathcal{PSR})} : M(p) \to \mathbb{R}$ *is continuous.*

The preceding result also implies that for any finite sample $X_1, \dots, X_n$, $f_n^{(\mathcal{SR})}$ (or $f_n^{(\mathcal{PSR})}$) is LSC (or continuous, respectively). Lemma 4.7 plays an important role in developing theoretical properties of SR and PSR means, which we present in the subsequent sections.

4.2. Existence of scaling-rotation means

The SR mean set $E^{(\mathcal{SR})}$ consists of the minimizers of the function $f^{(\mathcal{SR})}$. To prove existence of SR means (or, equivalently, non-emptiness of $E^{(\mathcal{SR})}$) we use the fact that any LSC function on a compact set attains a minimum. For this purpose, we first verify *coercivity* of $f^{(\mathcal{SR})}$ (and $f^{(\mathcal{PSR})}$).

Proposition 4.8. *Let* P *be a Borel probability measure on* $\mathrm{Sym}^+(p)$*.*

(a) *There exists a compact set* $K \subset \mathrm{Sym}^+(p)$ *such that*

$$\inf_{S \in \mathrm{Sym}^+(p)} f^{(\mathcal{SR})}(S) = \inf_{S \in K} f^{(\mathcal{SR})}(S). \tag{4.4}$$

(b) *There exists a compact set* $\widetilde{K} \subset M(p)$ *such that*

$$\inf_{m \in M(p)} f^{(\mathcal{PSR})}(m) = \inf_{m \in \widetilde{K}} f^{(\mathcal{PSR})}(m). \tag{4.5}$$

Proposition 4.8 says that $f^{(\mathcal{SR})}$ (and $f^{(\mathcal{PSR})}$, respectively) is coercive, i.e. uniformly large outside some compact set, and, under the finite-variance condition, has a (non-strictly) smaller value somewhere inside that compact set. Using this fact and the lower semicontinuity of $f^{(\mathcal{SR})}$ (respectively, $f^{(\mathcal{PSR})}$), we show in Theorem 4.9 that the SR and PSR mean sets are non-empty. In this theorem, the bounded support condition for P is used only to ensure the lower semicontinuity of $f^{(\mathcal{SR})}$.

Theorem 4.9. *Let* P *be a Borel probability measure on* $\mathrm{Sym}^+(p)$*.*

(a) *If* P *is supported on a compact set, then* $E^{(\mathcal{SR})} \neq \varnothing$*.*
(b) *If* P *has finite PSR-variance, then* $E^{(\mathcal{PSR})} \neq \varnothing$*.*

Since the conditions of Theorem 4.9 are met for any empirical measure defined from a finite set $\{X_1, \dots, X_n\} \subset \mathrm{Sym}^+(p)$, a corollary of the population SR mean result is the existence of sample SR means:

Corollary 4.10. *For any finite n, and any $X_1,\dots,X_n \in \mathrm{Sym}^+(p)$, $E_n^{(\mathcal{SR})} \neq \varnothing$ and $E_n^{(\mathcal{PSR})} \neq \varnothing$.*

Remark 4.11. For any Borel-measurable $\mathrm{Sym}^+(p)$-valued random variable with finite PSR-variance, the PSR mean set $E^{(\mathcal{PSR})}$ is closed. In particular, every sample PSR mean set is closed. To verify this, recall from Lemma 4.7 that $f^{(\mathcal{PSR})}$ is continuous. The PSR mean set is a level set of a continuous function, and therefore is closed. Moreover, as seen in Proposition 4.8, the closed set $E^{(\mathcal{PSR})}$ is a subset of a compact set, thus is compact as well.

4.3. Uniqueness of PSR means

Much work has been done on the question of uniqueness of the Fréchet mean of Riemannian manifold-valued observations. It is known that the Fréchet mean is unique as long as the support of the probability distribution lies within a geodesic ball of a certain radius (see, for example, Afsari (2011)). Although $d_{\mathcal{SR}}$ is not a geodesic distance on $\mathrm{Sym}^+(p)$, we can obtain a similar result for a kind of uniqueness of the PSR mean.

For any $X \in \mathrm{Sym}^+(p)$, recall from (2.6) that $\mathcal{F}^{-1}(X) = \{h\cdot(UR,D) : R \in G_D, h \in \mathcal{G}(p)\}$ for an eigen-decomposition (U,D) of X. Since the finite group $\mathcal{G}(p)$ acts freely and isometrically on $M(p)$, for any $h \in \mathcal{G}(p)$ and $m \in M(p)$,

$$\begin{aligned} d_{\mathcal{PSR}}(X,m) &= \inf_{R\in G_D, h'\in\mathcal{G}(p)} d_M(h'\cdot(UR,D),m) \\ &= \inf_{R\in G_D, h\cdot h'\in\mathcal{G}(p)} d_M(h\cdot h'\cdot(UR,D),h\cdot m) = d_{\mathcal{PSR}}(X,h\cdot m). \end{aligned}$$

For a sample $X_1,\dots,X_n \in \mathrm{Sym}^+(p)$, we have thus

$$f_n^{(\mathcal{PSR})}(m) = \frac{1}{n}\sum_{i=1}^n d^2_{\mathcal{PSR}}(X_i,m) = \frac{1}{n}\sum_{i=1}^n d^2_{\mathcal{PSR}}(X_i,h\cdot m) = f_n^{(\mathcal{PSR})}(h\cdot m) \tag{4.6}$$

for any $h \in \mathcal{G}(p)$ and $m \in M(p)$. It follows from (4.6) that for any $m \in E_n^{(\mathcal{PSR})}$, the remaining members of its orbit $\mathcal{G}(p)\cdot m$ (see (2.7)) also belong to $E_n^{(\mathcal{PSR})}$. Thus, $E_n^{(\mathcal{PSR})}$ will contain at least $2^{p-1}p!$ elements. In the case where $E_n^{(\mathcal{PSR})}$ only contains $2^{p-1}p!$ elements (necessarily belonging to the same orbit), we will say that the sample PSR mean is *unique up to the action of* $\mathcal{G}(p)$. The notion of uniqueness (up to the action of $\mathcal{G}(p)$) for the population PSR mean in $E^{(\mathcal{PSR})}$ is defined similarly.

The following lemma yields a useful lower bound on the distance between distinct eigen-decompositions of an SPD matrix in S_p^{top}. (Note that for any $X \in S_p^{\mathrm{lwr}}$, the set of eigen-decompositions of X is not discrete, so two eigen-decompositions of X may be arbitrarily close to each other.)

Lemma 4.12.

(a) *For any $(U,D) \in M(p)$ and for any $h \in \mathcal{G}(p)\setminus\{I_p\}$,*

$$d_M((U,D),h\cdot(U,D)) \ge \sqrt{k}\beta_{\mathcal{G}(p)}$$

where $\beta_{\mathcal{G}(p)} := \min_{h\in\mathcal{G}(p)\setminus\{I_p\}} d_{SO}(I_p,h) = \min_{h\in\mathcal{G}(p)\setminus\{I_p\}} \frac{1}{\sqrt{2}}\|\mathrm{Log}(h)\|_F$.

(b) *The quantity $\beta_{\mathcal{G}(p)}$ satisfies $\beta_{\mathcal{G}(p)} \le \frac{\pi}{2}$ for any $p \ge 2$.*

(c) *For any $X \in S_p^{\mathrm{top}}$, any two distinct eigen-decompositions (U_X,D_X) and (U'_X,D'_X) of X satisfy $d_M((U_X,D_X),(U'_X,D'_X)) \ge \sqrt{k}\beta_{\mathcal{G}(p)}$.*

Remark 4.13. It is easily checked that $\beta_{\mathcal{G}(p)} = \frac{\pi}{2}$ when $p = 2,3$.

In Theorem 4.15 below, we provide a sufficient condition for uniqueness (up to the action of $\mathcal{G}(p)$) of the PSR means. In preparation, we first provide a sufficient condition for a distribution on $M(p)$ to have a unique Fréchet mean. Recall that $(M(p), g_M)$ is a Riemannian manifold, which in turn implies that $(M(p), d_M)$ is a metric space. The Fréchet mean set for a probability distribution P on $M(p)$ is thus well-defined.

Lemma 4.14. *Let $\tilde{P}$ be a Borel probability measure on $M(p)$. Suppose that* $\mathrm{supp}(\tilde{P})$*, the support of $\tilde{P}$, satisfies*

$$\mathrm{supp}(\tilde{P}) \subseteq B_r^{d_M}(m_0) \tag{4.7}$$

for some $r \le \sqrt{k}\beta_{\mathcal{G}(p)}$ and some $m_0 \in M(p)$. Then there exists a unique Fréchet mean $\bar{m}(\tilde{P}) := \mathrm{argmin}_{m \in M(p)} \int_{M(p)} d_M^2(\tilde{X}, m)\tilde{P}(d\tilde{X})$ of P, and $\bar{m}(\tilde{P}) \in B_r^{d_M}(m_0)$.

Similarly to Lemma 4.14, if a deterministic sample $m_1, \ldots, m_n \in M(p)$ lies in $B_r^{d_M}(m_0)$ $(i = 1, \ldots, n)$ for some $r \le \sqrt{k}\beta_{\mathcal{G}(p)}$ and $m_0 \in M(p)$, then the sample Fréchet mean $\bar{m} := \mathrm{argmin}_{m \in M(p)} \frac{1}{n}\sum_{i=1}^n d_M^2(m_i, m)$ is unique and lies in $B_r^{d_M}(m_0)$.

Theorem 4.15. *Suppose the probability measure P on* $\mathrm{Sym}^+(p)$ *is absolutely continuous with respect to volume measure and that for two independent* $\mathrm{Sym}^+(p)$*-valued random variables X_1, X_2 whose distribution is P,*

$$P(d_{\mathcal{SR}}(X_1, X_2) < r'_{cx}) = 1, \quad \text{where } r'_{cx} := \frac{\sqrt{k}\beta_{\mathcal{G}(p)}}{4}. \tag{4.8}$$

Then the population PSR mean set $E^{(\mathcal{PSR})}$ is unique up to the action of $\mathcal{G}(p)$.

The number r'_{cx} is a lower bound on the regular convexity radius of the quotient space $M(p)/\mathcal{G}(p)$ with the induced Riemannian structure, as shown in Groisser, Jung and Schwartzman (2023). This ensures that a ball in $M(p)/\mathcal{G}(p)$ with radius less than r'_{cx} is convex. The quotient space $M(p)/\mathcal{G}(p)$ "sits" between $M(p)$ and $\mathrm{Sym}^+(p)$; any $X \in S_p^{\mathrm{top}}$ coincides with an element in $M(p)/\mathcal{G}(p)$, but there are multiple (in fact, infinitely many) elements in $M(p)/\mathcal{G}(p)$ corresponding to any $X \in S_p^{\mathrm{lwr}}$ (cf. (2.6)). Lemma 4.12 shows that $r'_{cx} \le \sqrt{k}\pi/8$. In contrast, the regular convexity radius of $(M(p), g_M)$ is $\sqrt{k}\pi/2$, which is much larger than r'_{cx}. Even though we work with the eigen-decomposition space $M(p)$, in Theorem 4.15 we require data-support diameter at most $r'_{cx} < \sqrt{k}\pi/2$ since, if $d_{\mathcal{SR}}(S_1, S_2) \ge r'_{cx}$ for some $S_1, S_2 \in \mathrm{Sym}^+(p)$, then there may be two or more eigen-decompositions of S_1 that are both closest to an eigen-decomposition of S_2.

The assumption of absolute continuity of P in Theorem 4.15 enables us to restrict our attention to the probability-1 event for which the random variables lie in the top stratum S_p^{top} of $\mathrm{Sym}^+(p)$, since the complement of S_p^{top} has volume zero in $\mathrm{Sym}^+(p)$. Corollary 4.16 below explicitly states this restriction as a sufficient condition for the uniqueness of sample PSR means of a deterministic sample. We also show that the estimation procedure (Procedure 3.8) will yield the unique (up to the action of $\mathcal{G}(p)$) sample PSR mean.

Corollary 4.16. *Assume $X_1, \ldots, X_n \in S_p^{\mathrm{top}}$. If*

$$d_{\mathcal{SR}}(X_i, X_j) < r'_{cx} \tag{4.9}$$

for all $i, j = 1, \ldots, n$*, then*

(a) *the sample PSR mean is unique up to the action of* $\mathcal{G}(p)$*;*
(b) *choosing an eigen-decomposition of any observation from the sample as the initial guess will lead Procedure 3.8 to converge to the sample PSR mean after one iteration.*

Remark 4.17. The data-diameter condition (4.9) in Corollary 4.16 is satisfied under either of the following two conditions (in the presence of the assumption $X_i \in S_p^{\text{top}}$):

(i) There exists an $S_0 \in S_p^{\text{top}}$ such that $d_{\mathcal{SR}}(S_0, X_i) < r'_{cx}/2$ for all $i = 1, \ldots, n$.
(ii) There exists an $m \in M(p)$ such that $d_{\mathcal{PSR}}(X_i, m) < r'_{cx}/2$ for all $i = 1, \ldots, n$.

Similarly, the condition that $d_{\mathcal{SR}}(X_1, X_2) < r'_{cx}$ almost surely in Theorem 4.15 is guaranteed by either (i) or (ii) above, when the latter two conditions are modified probabilistically. In condition (i) above, it is necessary for the center of the open ball (data support) to lie in the top stratum, due to the fact that the functions $d_{\mathcal{SR}}(\cdot, X_i)$ are, in general, only LSC (not continuous) at points belonging to S_p^{lwr}. For an $S_0 \in S_p^{\text{lwr}}$, even if a condition $d_{\mathcal{SR}}(S_0, X_i) < \epsilon$ ($i = 1, \ldots, n$) is satisfied for arbitrarily small ϵ, $d_{\mathcal{SR}}(X_i, X_j)$ may be larger than r'_{cx}. Proof of the statements given in this remark can be found in the Supplementary Material (Jung et al., 2024).

If the data-support is small enough to satisfy (4.8) and also is far from the lower stratum (satisfying the conditions in Theorem 3.7), then the SR mean is unique, as the following corollary states.

Corollary 4.18. *Let X be a* $\text{Sym}^+(p)$*-valued random variable following the distribution P. Assume that there exist* $S_0 \in S_p^{\text{top}}$ *and* $r < \min\{\delta(S_0)/3, r'_{cx}/2\}$ *satisfying* $P(d_{\mathcal{SR}}(S_0, X_i) \le r) = 1$*. Then, (i) the PSR mean is unique up to the action of* $\mathcal{G}(p)$*, (ii)* $E^{(\mathcal{SR})} \subset S_p^{\text{top}}$*, and (iii)* $E^{(\mathcal{SR})} = \mathcal{F}(E^{(\mathcal{PSR})})$ *is a singleton set.*

4.4. Asymptotic properties of the sample PSR means

This subsection addresses two aspects of the asymptotic behavior of the sample PSR mean $E_n^{(\mathcal{PSR})}$: (i) strong consistency of $E_n^{(\mathcal{PSR})}$ with the population PSR mean set $E^{(\mathcal{PSR})}$ and (ii) the large-sample limiting distribution of a sample PSR mean. Much work has been done to establish consistency and central limit theorem-type results for sample Fréchet means on Riemannian manifolds and metric spaces (Bhattacharya and Patrangenaru (2003), Bhattacharya and Patrangenaru (2005), Bhattacharya and Lin (2017), Eltzner and Huckemann (2019)). Estimation of the PSR mean does not fit into the context of estimation on Riemannian manifolds or metric spaces since the sample space $\text{Sym}^+(p)$ and *parameter space* $M(p)$ are different. Moreover, as we have seen, the PSR means are never unique. With this in mind, we apply the framework of generalized Fréchet means on general product spaces in Huckemann (2011a) and Huckemann (2011b) to our PSR mean estimation context, enabling us to establish conditions for strong consistency and for a central limit theorem.

We now establish a strong-consistency result for $E_n^{\mathcal{PSR}}$. Throughout this subsection, let $X, X_1, \ldots$ be independent random variables mapping from a complete probability space $(\Omega, \mathcal{A}, \mathcal{P})$ to $\text{Sym}^+(p)$ equipped with its Borel σ-field, and let P be the induced Borel probability measure on $\text{Sym}^+(p)$. The sets $E^{(\mathcal{PSR})}$ and $E_n^{(\mathcal{PSR})}$ denote the population and sample PSR-mean sets defined by P and $X_1, \ldots, X_n$, respectively.

Theorem 4.19. *Assume that P has finite PSR-variance. Then*

$$\lim_{n\to\infty} \sup_{m\in E_n^{(\mathcal{PSR})}} d_M(m, E^{(\mathcal{PSR})}) = 0 \tag{4.10}$$

almost surely.

Our proof of Theorem 4.19 is contained in Section B.5 of the Supplementary Material (Jung et al., 2024). There, we closely follow the arguments of Huckemann (2011b) used in verifying the conditions required to establish strong consistency of the generalized Fréchet means. However, the theorems of Huckemann (2011b) are not directly applied since the function $d_{\mathcal{PSR}}$ is not continuous. Nevertheless, the Fréchet-type objective function $f^{(\mathcal{PSR})} : M(p) \to \mathbb{R}$ is continuous, as shown in Lemma 4.7, a fact that plays an important role in the proof of Theorem 4.19. Schötz (2022) extends the results of Huckemann (2011b) by, among other things and in our notation, allowing for $d_{\mathcal{PSR}}(X,\cdot)$ to be only LSC. However, this is not actually helpful for $d_{\mathcal{PSR}}$ either, because $d_{\mathcal{PSR}}$ is actually *continuous* with respect to the second variable (it is LSC with respect to the *first* variable); see Lemma 4.1 and Theorem 4.3.

In the proof of Theorem 4.19, we first show that with probability 1

$$\cap_{k=1}^{\infty} \overline{\cup_{n=k}^{\infty} E_n^{(\mathcal{PSR})}} \subset E^{(\mathcal{PSR})}. \tag{4.11}$$

In the terminology of Huckemann (2011b), (4.11) is called strong consistency of $E_n^{(\mathcal{PSR})}$ as an estimator of $E^{(\mathcal{PSR})}$ in the sense of Ziezold (1977). Our result (4.10) is equivalent to strong consistency in the sense of Bhattacharya and Patrangenaru (2003) (again using the terminology of Huckemann (2011b)), as shown in Lemma B.20 in the Supplementary Material (Jung et al., 2024). Schötz (2022) classified three types of convergence for a sequence of sets, referring to (4.11) as convergence *in the outer limit*, and to (4.10) as convergence *in the one-sided Hausdorff distance*. The last type of convergence is convergence *in Hausdorff distance*. Recall that for a metric space (M,d) the Hausdorff distance between non-empty sets $A,B \subset M$ is $d_H(A,B) := \max\{\sup_{m\in A} d(m,B), \sup_{m\in B} d(A,m)\}$.

Theorem 4.19 states that, with probability 1, any sequence $m_n \in E_n^{(\mathcal{PSR})}$ of sample PSR means will eventually lie in an arbitrarily small neighborhood of the population PSR mean set as the sample size n increases. But, conceivably there could be a population PSR mean in $E^{(\mathcal{PSR})}$ with no sample PSR mean nearby even for large n, in which case $d_H(E_n^{(\mathcal{PSR})}, E^{(\mathcal{PSR})})$ would not approach zero. In other words, $E_n^{(\mathcal{PSR})}$ would be a strongly consistent estimator of $E^{(\mathcal{PSR})}$ only with respect to the *one-sided* Hausdorff distance, not the (two-sided) Hausdorff distance. However, if the population PSR mean is unique up to the action of $\mathcal{G}(p)$, then $E_n^{(\mathcal{PSR})}$ *is* a strongly consistent estimator of $E^{(\mathcal{PSR})}$ with respect to Hausdorff distance on $(M(p), d_M)$, as shown next.

Corollary 4.20. *Assume that P has finite PSR-variance, and that $E^{(\mathcal{PSR})} = \mathcal{G}(p)\cdot\mu$ for some $\mu \in M(p)$. Then with probability 1,*

$$\lim_{n\to\infty} \sup_{m\in E^{(\mathcal{PSR})}} d_M(E_n^{(\mathcal{PSR})}, m) = 0 \tag{4.12}$$

and

$$\lim_{n\to\infty} d_H(E_n^{(\mathcal{PSR})}, E^{(\mathcal{PSR})}) = 0. \tag{4.13}$$

The strong consistency of sample PSR means with the population PSR means can be converted to strong consistency of sample PSR means with the *population SR means*, as follows. For $S \in \mathrm{Sym}^+(p)$ and a set $\mathcal{E} \subset \mathrm{Sym}^+(p)$, we define $d_{\mathcal{SR}}(S,\mathcal{E}) := \inf_{E\in\mathcal{E}} d_{\mathcal{SR}}(S,E)$.

Corollary 4.21. *Assume that P has finite PSR-variance. Then,*

(i) $\lim_{n\to\infty} \sup_{S\in\mathcal{F}(E_n^{(\mathcal{PSR})})} d_{\mathcal{SR}}(S,\mathcal{F}(E^{(\mathcal{PSR})})) = 0$ *almost surely.*

(ii) If, in addition, $E^{(\mathcal{SR})} \subset S_p^{\mathrm{top}}$, *then* $\lim_{n\to\infty} \sup_{S\in\mathcal{F}(E_n^{(\mathcal{PSR})})} d_{\mathcal{SR}}(S,E^{(\mathcal{SR})}) = 0$ *almost surely.*

(iii) If $E^{(\mathcal{SR})} \subset S_p^{\mathrm{top}}$ *and the population SR mean is unique with* $E^{(\mathcal{SR})} = \{\mu^{(\mathcal{SR})}\}$, *then* $\lim_{n\to\infty} d_{\mathcal{SR}}(\mathcal{F}(E_n^{(\mathcal{PSR})}),\mu^{(\mathcal{SR})}) = 0$ *almost surely.*

Note that in establishing a strong consistency property of $E_n^{(\mathcal{PSR})}$ with respect to population (partial) SR means, we assumed only that the *population* mean set $E^{(\mathcal{PSR})}$ is unique up to the action of $\mathcal{G}(p)$, not that the *sample* mean sets $E_n^{(\mathcal{PSR})}$ have this uniqueness property. We also did not assume that $E_n^{(\mathcal{PSR})} \subset M_p^{\mathrm{top}}$.

We next establish a central limit theorem for our estimator $E_n^{(\mathcal{PSR})}$. Our strategy is to closely follow the arguments in Bhattacharya and Lin (2017), Bhattacharya and Patrangenaru (2005), Eltzner et al. (2021), Huckemann (2011a), for deriving central limit theorems for (generalized) Fréchet means on a Riemannian manifold. In particular, our central limit theorem is expressed in terms of charts and the asymptotic distributions of "linearized" estimators.

Our parameter space of interest $M(p) = SO(p) \times \mathrm{Diag}^+(p)$ is a Riemannian manifold of dimension $d := \frac{(p-1)p}{2} + p$. As defined in Section 2.1, the tangent space at $(U,D) \in M(p)$ is $T_{(U,D)}M(p) = \{(AU,LD) : A \in \mathfrak{so}(p), L \in \mathrm{Diag}(p)\}$, which can be canonically identified with $\mathfrak{so}(p) \oplus \mathrm{Diag}(p)$, a vector space of dimension d.

At $(U,D) \in M(p)$, we use the local chart $(\mathcal{U}_{(U,D)}, \tilde{\varphi}_{(U,D)})$, where

$$\mathcal{U}_{(U,D)} = \{(V,\Lambda) \in M(p) : \|\mathrm{Log}(VU^T)\|_F < \pi\},$$

and where $\tilde{\varphi}_{(U,D)} : \mathcal{U}_{(U,D)} \to \mathfrak{so}(p) \oplus \mathrm{Diag}(p)$ is defined by

$$\tilde{\varphi}_{(U,D)}(V,\Lambda) = (\mathrm{Log}(VU^T), \mathrm{Log}(\Lambda D^{-1})). \tag{4.14}$$

Observe that $\tilde{\varphi}^{-1}_{(U,D)}(A,L) = (\mathrm{Exp}(A)U, \mathrm{Exp}(L)D)$. The maps $\tilde{\varphi}_{(U,D)}$ and $\tilde{\varphi}^{-1}_{(U,D)}$ are the Riemannian logarithm and exponential maps to (and from) the tangent space $T_{(U,D)}M(p)$, composed with the right-translation isomorphism between $T_{(U,D)}M(p)$ and $T_{(I,I)}M(p) = \mathfrak{so}(p) \oplus \mathrm{Diag}(p)$.

We also write the elements of $\mathfrak{so}(p) \oplus \mathrm{Diag}(p)$ in a coordinate-wise vector form. For each $(A,L) \in \mathfrak{so}(p) \oplus \mathrm{Diag}(p)$, define a suitable vectorization operator vec,

$$\mathrm{vec}(A,L) := \begin{pmatrix} \sqrt{k}\, x_{SO}(A) \\ x_{\mathcal{D}}(L) \end{pmatrix} \in \mathbb{R}^d, \tag{4.15}$$

where $x_{SO}(A) \in \mathbb{R}^{(p-1)p/2}$ consists of the upper triangular entries of A (in the lexicographical ordering) and $x_{\mathcal{D}}(L) = (L_{11},\dots,L_{pp})^T \in \mathbb{R}^p$ consists of the diagonal entries of L. The inverse vectorization operator vec^{-1} is well-defined as well. We use the notation $\phi_{(U,D)}(\cdot,\cdot) := \mathrm{vec} \circ \tilde{\varphi}_{(U,D)}(\cdot,\cdot)$ and $\phi^{-1}_{(U,D)}(\cdot) := \tilde{\varphi}^{-1}_{(U,D)} \circ \mathrm{vec}^{-1}(\cdot)$.

Assume the following.

(A1) The probability measure P induced by X on $\mathrm{Sym}^+(p)$ is absolutely continuous with respect to volume measure, and has finite PSR-variance.

(A2) $E^{(\mathcal{PSR})}$ is unique up to the action of $\mathcal{G}(p)$. With probability 1, so is $E_n^{(\mathcal{PSR})}$ (for every n).
(A3) For some $m_0 \in E^{(\mathcal{PSR})}$, $P(d_{\mathcal{PSR}}(X, m_0) < r'_{cx}) = 1$.

The absolute continuity assumption (A1) ensures that any volume-zero (Lebesgue-measurable) subset of $\mathrm{Sym}^+(p)$ has probability zero. In particular, $P(X \in S_p^{\mathrm{top}}) = 1 - P(X \in S_p^{\mathrm{lwr}}) = 1$. This fact greatly simplifies our theoretical development.

The uniqueness assumption (A2) ensures that $E_n^{(\mathcal{PSR})}$ converges almost surely to $E^{(\mathcal{PSR})}$ with respect to the Hausdorff distance (by Corollary 4.20). Therefore, for any $m_0 \in E^{(\mathcal{PSR})}$, there exists a sequence $m_n \in E_n^{(\mathcal{PSR})}$ satisfying $d_M(m_n, m_0) \to 0$ (or, equivalently, $\phi_{m_0}(m_n) \to \phi_{m_0}(m_0) = 0$) as $n \to \infty$ almost surely. Assumption (A2) also guarantees that if (A3) is true for some PSR mean $m_0 \in E^{(\mathcal{PSR})}$ then it is true for any other PSR mean in $E^{(\mathcal{PSR})}$.

The radius $r'_{cx} = \sqrt{k}\beta_{\mathcal{G}(p)}/4$ in Assumption (A3) previously appeared in Theorem 4.16, where the bounded-support assumption was used to ensure uniqueness of one element of a minimal pair (see Definition 2.2) when the other element is fixed. Similarly, assumptions (A1) and (A3) ensure that, with probability 1, for each X_i there exists a unique $m_i \in \mathcal{F}^{-1}(X_i)$ such that $m_i \in B^{d_M}_{r'_{cx}}(m_0)$, a radius-$r'_{cx}$ ball in $M(p)$ centered at m. A stronger version of this fact will be used (in the proof of Theorem 4.22, to be given shortly) to rewrite the objective function $f_n^{(\mathcal{PSR})}$ involving $d_{\mathcal{PSR}}$ as a Fréchet objective function $m \mapsto \frac{1}{n}\sum_{i=1}^n d_M^2(m_i, m)$, with probability 1.

In addition, the bounded support condition (A3) ensures that with probability 1 the function $d^2_{\mathcal{PSR}}(X, \cdot)$ is smooth (C^∞) and convex on a convex set. Using this fact and geometric results given in Afsari (2011) and Afsari, Tron and Vidal (2013), we show in the proof that the gradient

$$\mathrm{grad}_x\, d^2_{\mathcal{PSR}}(X, \phi^{-1}_{m_0}(x)) := \left(\frac{\partial}{\partial x_i} d^2_{\mathcal{PSR}}(X, \phi^{-1}_{m_0}(x))\right)_{i=1,\dots,d}$$

at $x = 0$ has mean zero and has a finite covariance matrix $\Sigma_P := \mathrm{Cov}(\mathrm{grad}_x\, d^2_{\mathcal{PSR}}(X, \phi^{-1}_{m_0}(0)))$. (We conjecture that (A1) guarantees that Σ_P is also positive-definite.) Likewise, as we will see in the proof, the differentiability and (strict) convexity of $d^2_{\mathcal{PSR}}(X, \phi^{-1}_{m_0}(\cdot))$ guarantee that the expectation of the Hessian $H_P(x) := E\left(\mathbf{H}d^2_{\mathcal{PSR}}(X, \phi^{-1}_{m_0}(x))\right)$ exists and is positive definite at $x = 0$. Write $H_P := H_P(0)$.

In summary, Assumptions (A1)—(A3) enable us to use a second-order Taylor expansion for $f_n^{(\mathcal{PSR})}$, to which the classical central limit theorem and the law of large numbers are applied. Such an approach was used in Bhattacharya and Patrangenaru (2005) and Huckemann (2011a). In particular, our proof for part (b) of Theorem 4.22 (in Section B.5.2 of the Supplementary Material (Jung et al., 2024)) closely follows the proof of Theorem 6 of Huckemann (2011a).

Theorem 4.22. *Suppose that Assumptions (A1)—(A3) are satisfied, and let $m_0 \in E^{(\mathcal{PSR})}$ be a PSR mean. Let $\{m'_n \in E_n^{(\mathcal{PSR})}\}$ be any choice of sample PSR mean sequence. For each n, let $m_n \in \mathrm{argmin}_{m \in \mathcal{G}(p)\cdot m'_n} d_M(m, m_0)$. Then, with probability 1, the sequence $\{m_n\}$ is determined uniquely. Furthermore,*

(a) *$m_n \to m_0$ almost surely as $n \to \infty$, and*
(b) *$\sqrt{n}\phi_{m_0}(m_n) \to N_d(0, H_P^{-1}\Sigma_P H_P^{-1})$ in distribution as $n \to \infty$.*

Estimating the covariance matrix ($H_P^{-1}\Sigma_P H_P^{-1}$ in our case) of the limiting Gaussian distribution (for Riemannian manifold-valued Fréchet means) is a difficult task. For general Riemannian manifold-valued Fréchet means, Bhattacharya and Patrangenaru (2005) and Bhattacharya and Bhattacharya (2012) suggest using a moment estimator for H_P and Σ_P. This however requires specifying the second derivatives of $d^2_{\mathcal{PSR}}(X, \phi_{m_0}(\cdot))$. We note that in the literature, explicit expressions for H_P and Σ_P

are only available for geometrically very simple manifolds, with a high degree of symmetry, such as spheres. As an instance, see Hotz and Huckemann (2015) and Section 5.3 of Bhattacharya and Bhattacharya (2012) for the cases where the data and the Fréchet mean lie in the unit circle S^1 and the more general unit sphere S^d, respectively. Others, including Eltzner and Huckemann (2019), simply use the sample covariance matrix of $\{\phi_{m_n}(m_{X_i}) : i = 1,\dots,n\}$ (in our notation) as an estimate of $H_P^{-1}\Sigma_P H_P^{-1}$. In Section 5, we will use a bootstrap estimator of $\mathrm{Var}(\phi_{m_0}(m_n))$, the variance of $\phi_{m_0}(m_n)$, instead of directly estimating $H_P^{-1}\Sigma_P H_P^{-1}$. Out bootstrap estimator is defined as follows.

Choose a PSR mean m_n computed from the original sample $\{X_1,\dots,X_n\}$. For the bth bootstrap sample (that is, a simple random sample of size n from the set $\{X_1,\dots,X_n\}$, treated as a fixed set, with replacement), let m_b^* be the PSR mean of the bootstrap sample that is closest to m_n. (For the purpose of defining the bootstrap estimator, we are assuming that such an m_b^* is unique.) The bootstrap estimator of $\mathrm{Var}(\phi_{m_0}(m_n))$ is then defined to be

$$\widehat{\mathrm{Var}}_{\mathrm{boot}}(\phi_{m_0}(m_n)) := \frac{1}{B}\sum_{b=1}^{B}\phi_{m_n}(m_b^*)\cdot(\phi_{m_n}(m_b^*))^T,$$

where B is the number of bootstrap replicates.

5. Numerical examples

5.1. Comparison with other geometric frameworks

In analyzing SPD matrices, the scaling-rotation (SR) framework has an advantage in interpretation as it allows describing the changes of SPD matrices in terms of rotation and scaling of the corresponding ellipsoids. For example, only the SR framework yields interpolation curves which consist of pure rotation when the endpoints differ only by rotation, when compared to the commonly used log-Euclidean (Arsigny et al., 2006/07), affine-invariant (Fletcher et al., 2004, Pennec, Fillard and Ayache, 2006), Procrustes (Dryden, Koloydenko and Zhou, 2009) and log-Cholesky (Lin, 2019) interpolation curves; see Section C.2 of the Supplementary Material (Jung et al., 2024) for examples regarding diffusion tensor imaging.

In this subsection, we illustrate situations under which averaging via the scaling-rotation framework has similar interpretive advantages over the affine-invariant mean. Comparisons to means obtained under log-Euclidean, Procrustes and log-Cholesky metrics are qualitatively similar to the conclusions below, and can be found in the Supplementary Material (Jung et al., 2024). The affine-invariant (AI) mean $\bar{X}^{(\mathrm{AI})}$ for a sample of SPD matrices $X_1,\dots,X_n \in \mathrm{Sym}^+(p)$ is the sample Fréchet mean with respect to the affine-invariant metric d_{AI}:

$$\bar{X}^{(\mathrm{AI})} = \operatorname*{argmin}_{M\in\mathrm{Sym}^+(p)}\sum_{i=1}^{n} d_{AI}^2(M,X_i),$$

where $d_{AI}(X,Y) = \|\mathrm{Log}(X^{-1/2}YX^{-1/2})\|_F$. The AI mean exists and is unique for any finite sample (Pennec, Fillard and Ayache, 2006).

A sample of size $n = 200$, sampled from a model detailed in Section C.3 of the Supplementary Material (Jung et al., 2024), is plotted in Fig. 1. There, we have used two different types of "linearizations" of $\mathrm{Sym}^+(2)$, as explained below.

The *Log-Euclidean coordinates* on $\mathrm{Sym}^+(2)$ are given by the three free parameters y_{11}, y_{22} and $\sqrt{2}y_{12}$ of $Y = (y_{ij}) = \mathrm{Log}(X) \in \mathrm{Sym}(2)$. Write $\mathrm{vecd}(Y) := (y_{11}, y_{22}, \sqrt{2}y_{12})^T \in \mathbb{R}^3$. These coordinates

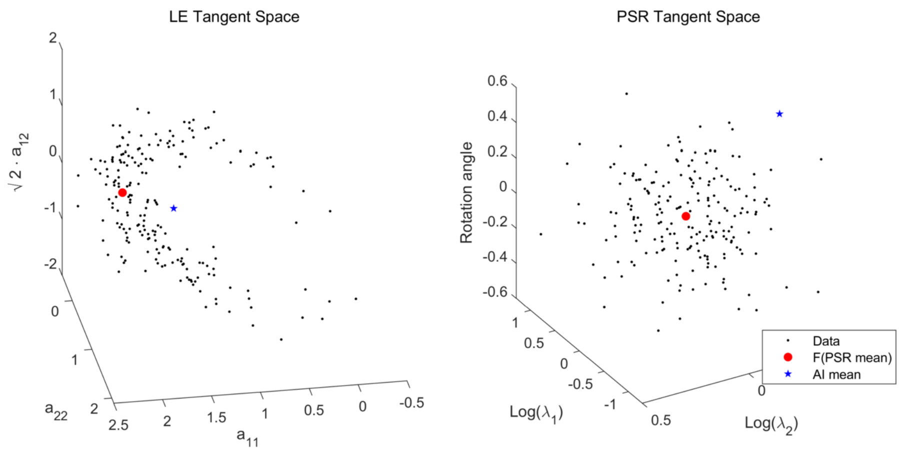

Figure 1. A sample of SPD matrices shown in the Log-Euclidean (LE) coordinates (left) and the PSR coordinates (right), overlaid with the PSR mean and AI mean. For this data set, PSR mean appears to be a better representative for the data, while the AI mean does not lie in the data-dense region. See Section 5.1 for details.

are chosen so that for any two vectors $(\text{vecd}(X), \text{vecd}(Y)) = (x, y)$, the usual inner product $\langle x, y\rangle = x^T y$ corresponds to the Riemannian metric when $X, Y \in \text{Sym}(2)$ are viewed as tangent vectors in the affine-invariant framework. The left panel of Fig. 1 plots the data on the Log-Euclidean coordinates.

The *PSR coordinates*, used in the right panel of the figure for the same data, come from the coordinates defined on a tangent space of the eigen-decomposition space $M(p)$. More precisely, given a reference point $(U, D) \in M(p)$, we use the local chart $(\mathcal{U}_{(U,D)}, \phi_{(U,D)})$ defined in (4.14), followed by the vectorization via vec (see (4.15)), to determine a coordinate system. To illustrate this concretely, let $p = 2$. Then $\tilde{\varphi}_{(U,D)}(V, \Lambda) = (\text{Log}(VU^T), \text{Log}(\Lambda D^{-1})) =: (A, L) \in \mathfrak{so}(p) \oplus \text{Diag}(p)$. The first coordinate of $x_{V,\Lambda} := \text{vec}(\phi_{(U,D)}(V, \Lambda)) \in \mathbb{R}^3$ is the free parameter a_{21} of A (multiplied by the scale parameter $\sqrt{k}$), and corresponds to the rotation angle of VU^T in radians (scaled by $\sqrt{k}$). The second and last coordinates of $x_{V,\Lambda}$ are the diagonal entries of L. Multiplying by $\sqrt{k}$ as above affords us the convenience that for any two x, y, the usual inner product $\langle x, y\rangle = x^T y$ corresponds to the Riemannian metric g_M we have assumed on the tangent spaces of $M(p)$.

When representing SPD-valued data $X_1, \ldots, X_n \in \text{Sym}^+(2)$ in PSR coordinates, we choose the reference point (U, D) to be an arbitrarily chosen PSR mean $\hat{m}^{\mathcal{PSR}}$ of the data. Care is needed since there are multiple eigen-decompositions corresponding to each observation X_i. For each X_i, an eigen-decomposition $m_i \in \mathcal{F}^{-1}(X_i) \subset M(p)$ is chosen so that m_i has the smallest geodesic distance from $\hat{m}^{\mathcal{PSR}}$ among all elements of $\mathcal{F}^{-1}(X_i)$. The right panel of Fig. 1 plots the same data as in the left panel, but in these PSR coordinates.

The AI mean and a PSR mean for this data set are also plotted in Fig. 1. (The log-Euclidean mean is located close to the AI mean, and is omitted from the figure.) It can be seen that major modes of variation in the data are well described in terms of rotation angles and scaling, while the variation appears to be highly non-linear in Log-Euclidean (LE) coordinates. As one might expect from this non-linearity, we observe that the AI mean is located far from the data, while the PSR mean appears to be a better representative of the data.

In the opposite direction, we also considered a data set sampled from an SPD-matrix log-normal distribution (Schwartzman, 2016). Note that the SPD-matrix log-normal distributions on $\text{Sym}^+(p)$ correspond to a multivariate normal distribution in Log-Euclidean coordinates. The data and their AI and

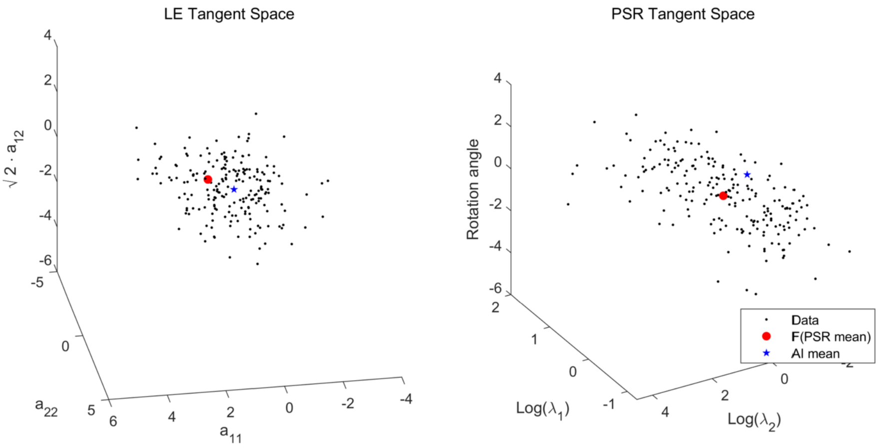

Figure 2. A sample of SPD matrices shown in the Log-Euclidean (LE) coordinates (left) and the PSR coordinates (right), overlaid with the PSR mean and AI mean. See Section 5.1 for details.

PSR means are plotted in Fig. 2. While the AI mean is well approximated by the average of data in LE coordinates, the PSR mean (in LE coordinates) is also not far from this average. Similarly, the PSR mean is approximately the average in PSR coordinates, and the AI mean is also not far. Therefore, we may conclude that using the SR framework and PSR means is beneficial especially if variability in the sample (or in a population) is pronounced in terms of rotations, while the cost of using the SR framework is small for the log-normal case.

5.2. An application to multivariate tensor-based morphometry

In Paquette et al. (2017), the authors compared the lateral ventricular structure in the brains of 17 pre-term and 19 full-term infant children. After an MRI scan of a subject's brain was obtained and processed through an image processing pipeline, the shape data collected at 102816 vertices on the surfaces of their left and right ventricles were mapped onto the left and right ventricles of a template brain image, after which the 2×2 Jacobian matrix J from that surface registration transformation was computed at each vertex for each subject. The deformation tensor $X = (J^T J)^{1/2} \in \mathrm{Sym}^+(2)$ was then computed at each vertex for each subject. To summarize the structure of the data, there are 102816 vertices along the surfaces of the template ventricles, and at each vertex there are deformation tensors (2×2 SPD matrices) from $n_1 = 17$ pre-term and $n_2 = 19$ full-term infants. We will call these group 1 and group 2, respectively.

One way that the authors tested for differences in ventricular shape between the two groups was by performing two-sample location tests at each vertex via use of the log-Euclidean version of Hotelling's T^2 test statistic introduced in Lepore et al. (2008). The log-Euclidean (LE) version of the T^2 test statistic is the squared Mahalanobis distance between the full-term and pre-term log-Euclidean sample means on the LE coordinates (defined in Section 5.1).

Similarly, one could also measure separation between groups by comparing their respective PSR means in PSR coordinates. For this two-group context, the reference point for the PSR coordinates is given by a PSR mean computed from pooled sample (with sample size $n_1 + n_2$).

We have chosen vertex 75412 as an example to illustrate a scenario in which two groups have little separation in the LE coordinates but are well-separated in the PSR coordinates. In the top row of

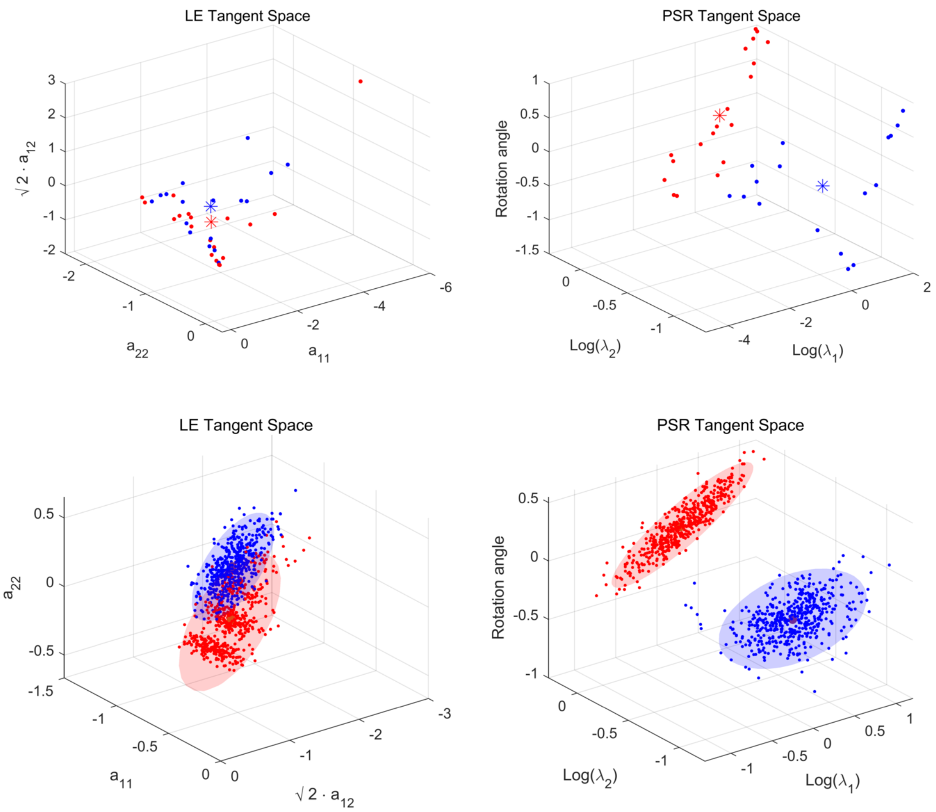

Figure 3. Real data example. (Top row) Observations corresponding to Group 1 (and Group 2) are shown in blue (and red, respectively) dots. The group-wise LE and PSR means are shown as the asterisks. (Bottom row) Bootstrap replications of the LE and PSR means (left and right panels, respectively) for each group, with the 95% approximate confidence regions shown as transparent ellipsoids. See Section 5.2 for details.

Figure 3, tensors from the two groups as well as the group-wise LE and PSR means are plotted in their respective coordinates. There is little visible separation between the two groups in the LE coordinates, while there is near-total separation in the PSR coordinates.

To visualize the sampling distributions of the group-wise means under the log-Euclidean and scaling-rotation frameworks, we computed 500 bootstrap sample means for each group, under both geometric frameworks. These are plotted in their respective tangent spaces in the bottom row of Figure 3. (See also Figure C.11 in the Supplementary Material (Jung et al., 2024), in which one can see that the (bootstrap) sampling distributions of the group-wise PSR means are approximately normal.) The nonparametric bootstrap provides an estimate of standard errors of the sample means, from which (bootstrap-approximated) parametric 95% confidence regions are obtained. For this, we assume normality, as suggested by the central limit theorem, Theorem 4.22, and obtain an approximate 95% confidence region given by $\{x \in \mathbb{R}^3 : x^T \hat{\Sigma}^{-1} x \leq \chi^2_{0.05,2}\}$, for each sample mean. Here, $\hat{\Sigma}$ is the sample covariance matrix of the bootstrap (group-wise LE or PSR) sample means, and $\chi^2_{0.05,2}$ is the 95% quantile of the χ^2_2 distribution. The resulting confidence regions are overlaid in the bottom row of Figure 3 as well. As in the top row, there is considerable overlap between the LE confidence regions, while there is complete separation between the two confidence regions for the group-wise PSR sample means, especially along the direction of rotation angles. We further utilize the asymptotic normality of the group-wise sample averages on the tangent space by conducting a approximate T^2 test for testing the null hypothesis of $\mu_1 = \mu_2$, where μ_i stands for the population mean vector of the ith group, against the general alternative $\mu_1 \neq \mu_2$. We used the unequal-covariance-variant of Hotelling's T^2 statistics and the approximate null distribution of Krishnamoorthy and Yu (2004) to obtain the p-value for this test. (See the Supplementary Material (Jung et al., 2024) for details.) When the LE means and LE coordinates are used, the p-value of the test is 0.3447. That is, there is not enough evidence to reject the null hypothesis when LE means are used. On the other hand, when PSR means are used, the corresponding p-value is nearly zero, reflecting the complete separation of confidence regions, seen in the bottom right panel of Figure 3. This example suggests that the scaling-rotation framework may be better at detecting group differences than other frameworks when most of the variability between the groups is due to rotation.

6. Discussion

We have presented the first statistical estimation methods for $\mathrm{Sym}^+(p)$ based on the scaling-rotation framework of Jung, Schwartzman and Groisser (2015). These estimation methods are intended to set the foundation for the development of scaling-rotation-framework-based statistical methods, such as testing the equality of two or more PSR means, testing for a variety of eigenvalue and eigenvector patterns of SPD matrices, and an analogue of principal component analysis for SPD-valued data. The scaling-rotation framework should also be particularly useful for diffusion tensor processing since the eigenvectors and eigenvalues of a diffusion tensor model the principal directions and intensities of water diffusion at a given voxel, and are thus the primary objects of interest.

The scaling-rotation estimation procedure presented here works well for $p = 2,3$. For these low-dimensional cases, the computation times for PSR means are significantly less than those for the AI means. However, for larger p, the number of eigen-decompositions of an SPD matrix from S_p^{top} grows rapidly with p, and the computational complexity of PSR mean computation increases. One avenue for interesting future work is to develop computational procedures for higher p. Another avenue is to develop a proper two-sample or multi-sample testing framework, and dimension-reduction and regression methods using the eigen-decomposition spaces, and to establish asymptotic and non-asymptotic properties of these statistical methods, reflecting the structure of $\mathrm{Sym}^+(p)$ as a stratified space under eigen-decomposition. Lastly, the authors are currently studying a genuine metric for $\mathrm{Sym}^+(p)$ that incorporates the benefits of the scaling-rotation framework. This metric can be used to analyze $\mathrm{Sym}^+(p)$-valued data within the framework of metric statistics (Dubey, Chen and Müller, 2024, Liu, Wang and Zhu, 2022).

Funding

The first author was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (No. 2019R1A2C2002256, RS-2023-00301976 and RS-2024-00333399).

Supplementary Material

Technical details and examples (DOI: 10.3150/24-BEJ1783SUPP; .pdf). This supplementary material contains technical details and proofs for all theoretical results in the article, as well as additional numerical examples.

References

Afsari, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346 https://doi.org/10.1090/S0002-9939-2010-10541-5

Afsari, B., Tron, R. and Vidal, R. (2013). On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM J. Control Optim.* **51** 2230–2260. MR3057324 https://doi.org/10.1137/12086282X

Alexander, D.C. (2005). Multiple-fiber reconstruction algorithms for diffusion MRI. *Ann. N.Y. Acad. Sci.* **1064** 113–133. https://doi.org/10.1196/annals.1340.018

Arsigny, V., Fillard, P., Pennec, X. and Ayache, N. (2006/07). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **29** 328–347. MR2288028 https://doi.org/10.1137/050637996

Batchelor, P.G., Moakher, M., Atkinson, D., Calamante, F. and Connelly, A. (2005). A rigorous framework for diffusion tensor calculus. *Magn. Reson. Med.* **53** 221–225. https://doi.org/10.1002/mrm.20334

Bhattacharya, A. and Bhattacharya, R. (2012). *Nonparametric Inference on Manifolds*. *Institute of Mathematical Statistics (IMS) Monographs* **2**. Cambridge: Cambridge Univ. Press. With applications to shape spaces. MR2934285 https://doi.org/10.1017/CBO9781139094764

Bhattacharya, R. and Lin, L. (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proc. Amer. Math. Soc.* **145** 413–428. MR3565392 https://doi.org/10.1090/proc/13216

Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498 https://doi.org/10.1214/aos/1046294456

Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634 https://doi.org/10.1214/009053605000000093

Carmichael, O., Chen, J., Paul, D. and Peng, J. (2013). Diffusion tensor smoothing through weighted Karcher means. *Electron. J. Stat.* **7** 1913–1956. MR3084676 https://doi.org/10.1214/13-EJS825

Collard, A., Bonnabel, S., Phillips, C. and Sepulchre, R. (2014). Anisotropy preserving DTI processing. *Int. J. Comput. Vis.* **107** 58–74.

Dryden, I.L., Koloydenko, A. and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** 1102–1123. MR2750388 https://doi.org/10.1214/09-AOAS249

Dubey, P., Chen, Y. and Müller, H.-G. (2024). Metric statistics: Exploration and inference for random objects with distance profiles. *Ann. Statist.* **52** 757–792. MR4744195 https://doi.org/10.1214/24-aos2368

Eltzner, B. and Huckemann, S.F. (2019). A smeary central limit theorem for manifolds with application to high-dimensional spheres. *Ann. Statist.* **47** 3360–3381. MR4025745 https://doi.org/10.1214/18-AOS1781

Eltzner, B., Galaz-García, F., Huckemann, S. and Tuschmann, W. (2021). Stability of the cut locus and a central limit theorem for Fréchet means of Riemannian manifolds. *Proc. Amer. Math. Soc.* **149** 3947–3963. MR4291592 https://doi.org/10.1090/proc/15429

Feragen, A. and Fuster, A. (2017). Geometries and interpolations for symmetric positive definite matrices. In *Modeling, Analysis, and Visualization of Anisotropy*(T. Schultz, E. Özarslan and I. Hotz, eds.). *Math. Vis.* 85–113. Cham: Springer. MR3822438

Feragen, A. and Nye, T. (2020). Statistics on stratified spaces. In *Riemannian Geometric Statistics in Medical Image Analysis* (X. Pennec, S. Sommer and T. Fletcher, eds.) 299–342. London: Elsevier/Academic Press. MR4254219 https://doi.org/10.1016/B978-0-12-814725-2.00016-9

Fletcher, P.T., Lu, C., Pizer, S.M. and Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imag.* **23** 995–1005.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. MR0027464

Gower, J.C. (1975). Generalized Procrustes analysis. *Psychometrika* **40** 33–51. MR0405725 https://doi.org/10.1007/BF02291478

Groisser, D., Jung, S. and Schwartzman, A. (2017). Geometric foundations for scaling-rotation statistics on symmetric positive definite matrices: Minimal smooth scaling-rotation curves in low dimensions. *Electron. J. Stat.* **11** 1092–1159. MR3631822 https://doi.org/10.1214/17-EJS1250

Groisser, D., Jung, S. and Schwartzman, A. (2021). Uniqueness questions in a scaling-rotation geometry on the space of symmetric positive-definite matrices. *Differential Geom. Appl.* **79** Paper No. 101798, 40. MR4296196 https://doi.org/10.1016/j.difgeo.2021.101798

Groisser, D., Jung, S. and Schwartzman, A. (2023). A genericity property of Fréchet sample means on Riemannian manifolds. Preprint. Available at arXiv:2309.13823.

Hotz, T. and Huckemann, S. (2015). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Ann. Inst. Statist. Math.* **67** 177–193. MR3297863 https://doi.org/10.1007/s10463-013-0444-7

Huckemann, S. (2011a). Inference on 3D Procrustes means: Tree bole growth, rank deficient diffusion tensors and perturbation models. *Scand. J. Stat.* **38** 424–446. MR2833839 https://doi.org/10.1111/j.1467-9469.2010.00724.x

Huckemann, S.F. (2011b). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.* **39** 1098–1124. MR2816349 https://doi.org/10.1214/10-AOS862

Huckemann, S.F. and Eltzner, B. (2021). Data analysis on nonstandard spaces. *Wiley Interdiscip. Rev.: Comput. Stat.* **13** Paper No. e1526, 19. MR4242812 https://doi.org/10.1002/wics.1526

Jung, S., Schwartzman, A. and Groisser, D. (2015). Scaling-rotation distance and interpolation of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **36** 1180–1201. MR3379023 https://doi.org/10.1137/140967040

Jung, S., Rooks, B., Groisser, D. and Schwartzman, A. (2024). Supplement to "Averaging symmetric positive-definite matrices on the space of eigen-decompositions." https://doi.org/10.3150/24-BEJ1783SUPP

Krishnamoorthy, K. and Yu, J. (2004). Modified Nel and van der Merwe test for the multivariate Behrens-Fisher problem. *Statist. Probab. Lett.* **66** 161–169. MR2029731 https://doi.org/10.1016/j.spl.2003.10.012

Lepore, F., Brun, C., Chou, Y.-Y., Chiang, M.-C., Dutton, R., Hayashi, K., Luders, E., Lopez, O., Aizenstein, H., Toga, A.W., Becker, J. and Thompson, P. (2008). Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. *IEEE Trans. Med. Imag.* **27** 129–141.

Lin, Z. (2019). Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM J. Matrix Anal. Appl.* **40** 1353–1370. MR4032859 https://doi.org/10.1137/18M1221084

Liu, H., Wang, X. and Zhu, J. (2022). Quantiles, ranks and signs in metric spaces. Preprint. Available at arXiv:2209.04090.

Manton, J.H. (2004). A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups. In *Int. Conf. Control Autom. Robot. Vis.* **3** 2211–2216.

Masarotto, V., Panaretos, V.M. and Zemel, Y. (2019). Procrustes metrics on covariance operators and optimal transportation of Gaussian processes. *Sankhya A* **81** 172–213. MR3982195 https://doi.org/10.1007/s13171-018-0130-1

Paquette, N., Shi, J., Wang, Y., Lao, Y., Ceschin, R., Nelson, M., Panigrahy, A. and Lepore, N. (2017). Ventricular shape and relative position abnormalities in preterm neonates. *NeuroImage Clin.* **15** 483–493.

Pennec, X., Fillard, P. and Ayache, N. (2006). A Riemannian framework for tensor computing. *Int. J. Comput. Vis.* **66** 41–66.

Schötz, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electron. J. Stat.* **13** 4280–4345. MR4023955 https://doi.org/10.1214/19-EJS1618

Schötz, C. (2022). Strong laws of large numbers for generalizations of Fréchet mean sets. *Statistics* **56** 34–52. MR4392146 https://doi.org/10.1080/02331888.2022.2032063

Schwartzman, A. (2016). Lognormal distributions and geometric averages of symmetric positive definite matrices. *Int. Stat. Rev.* **84** 456–486. MR3580425 https://doi.org/10.1111/insr.12113

Schwartzman, A., Dougherty, R.F. and Taylor, J.E. (2010). Group comparison of eigenvalues and eigenvectors of diffusion tensors. *J. Amer. Statist. Assoc.* **105** 588–599. MR2724844 https://doi.org/10.1198/jasa.2010.ap07291

Wang, Y., Salehian, H., Cheng, G. and Vemuri, B.C. (2014). Tracking on the product manifold of shape and orientation for tractography from diffusion MRI. *Conf Comput Vis Pattern Recognit Workshops* **2014** 3051–3056. https://doi.org/10.1109/CVPR.2014.390

Ziezold, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A* 591–602. Dordrecht-Boston, Mass.: Reidel. MR0501230

Received October 2023 and revised May 2024