SIAM J. Matrix Anal. Appl.
Vol. 36, No. 3, pp. 1180–1201

SCALING-ROTATION DISTANCE AND INTERPOLATION OF SYMMETRIC POSITIVE-DEFINITE MATRICES*

SUNGKYU JUNG†, ARMIN SCHWARTZMAN‡, AND DAVID GROISSER§

Abstract. We introduce a new geometric framework for the set of symmetric positive-definite (SPD) matrices, aimed at characterizing deformations of SPD matrices by individual scaling of eigenvalues and rotation of eigenvectors of the SPD matrices. To characterize the deformation, the eigenvalue-eigenvector decomposition is used to find alternative representations of SPD matrices and to form a Riemannian manifold so that scaling and rotations of SPD matrices are captured by geodesics on this manifold. The problems of nonunique eigen-decompositions and eigenvalue multiplicities are addressed by finding minimal-length geodesics, which gives rise to a distance and an interpolation method for SPD matrices. Computational procedures for evaluating the minimal scaling-rotation deformations and distances are provided for the most useful cases of 2×2 and 3×3 SPD matrices. In the new geometric framework, minimal scaling-rotation curves interpolate eigenvalues at constant logarithmic rate, and eigenvectors at constant angular rate. In the context of diffusion tensor imaging, this results in better behavior of the trace, determinant, and fractional anisotropy of interpolated SPD matrices in typical cases.

Key words. symmetric positive-definite matrices, eigen-decomposition, Riemannian distance, geodesics, diffusion tensors

AMS subject classifications. 15A18, 15A16, 53C20, 53C22, 57S15, 22E30

DOI. 10.1137/140967040

1. Introduction.

The analysis of symmetric positive-definite (SPD) matrices as data objects arises in many contexts. A prominent example is diffusion tensor imaging (DTI), which is a widely used technique that measures the diffusion of water molecules in a biological object [4, 16, 2]. The diffusion of water is characterized by a symmetric tensor that is represented by a 3×3 SPD matrix. SPD matrices also appear in other contexts, such as in tensor computing [24], in tensor-based morphometry [18], and as covariance matrices [30]. In recent years, statistical analyses of SPD matrices have received much attention [34, 26, 28, 27, 22, 33, 23].

The main challenge in the analysis of SPD matrices is that the set of $p \times p$ SPD matrices, $\mathrm{Sym}^+(p)$, is a proper open subset of a real matrix space, so it is not a vector space. This has led researchers to consider alternative geometric frameworks to handle analytic and statistical tasks for SPD matrices. The most popular framework is a Riemannian framework, where the set of SPD matrices is endowed with an affine-invariant Riemannian metric [21, 24, 17, 11]. The log-Euclidean metric, discussed in [3], is also widely used, because of its simplicity. In [10], Dryden, Koloydenko, and Zhou listed these popular approaches, including the Cholesky decomposition-based approach of [31] and their own approach, which they called the Procrustes distance. In [6], Bonnabel and Sepulchre proposed a different Riemannian approach for symmetric positive semidefinite matrices of fixed rank.

*Received by the editors April 30, 2014; accepted for publication (in revised form) by M. L. Overton May 29, 2015; published electronically August 11, 2015. This work was supported by NIH grant R21EB012177 and NSF grant DMS-1307178.

http://www.siam.org/journals/simax/36-3/96704.html

†Department of Statistics, University of Pittsburgh, Pittsburgh, PA 15260 (sungkyu@pitt.edu).

‡Department of Statistics, North Carolina State University, Raleigh, NC 27695 (aschwar@ncsu.edu).

§Department of Mathematics, University of Florida, Gainesville, FL 32611 (groisser@ufl.edu).

Although these approaches are powerful in generalizing statistics to SPD matrices, they are not easy to interpret in terms of SPD matrix deformations. In particular, in the context of DTI, tensor changes are naturally characterized by changes in diffusion orientation and intensity, but the above frameworks do not provide such an interpretation. In [25], Schwartzman proposed a scaling-rotation curve in $\mathrm{Sym}^+(p)$, which is interpretable as rotation of diffusion directions and scaling of the main modes of diffusivity. In this paper we develop a novel framework to formally characterize scaling-rotation deformations between SPD matrices and introduce a new distance, here called the scaling-rotation distance, defined by the minimum amount of rotation and scaling needed to deform one SPD matrix into another.

To this end, an alternative representation of $\mathrm{Sym}^+(p)$, obtained from the decomposition of each SPD matrix into an eigenvalue matrix and eigenvector matrix, is identified as a Riemannian manifold. This manifold, a generalized cylinder embedded in a higher-dimensional matrix space, is easy to endow with a Riemannian geometry. A careful analysis is provided to handle the case of equal eigenvalues and, more generally, the nonuniqueness of the eigen-decomposition. We show that the scaling-rotation curve corresponds to geodesics in the new geometry, and we characterize the family of geodesics. A minimal deformation of SPD matrices in terms of the smallest amount of scaling and rotation is then found by a minimal scaling-rotation curve, through a minimal-length geodesic. Sufficient conditions for the uniqueness of minimal curves are given.

The proposed framework not only provides a minimal deformation, it also yields a distance between SPD matrices. This distance function is a semimetric on $\mathrm{Sym}^+(p)$ and is invariant under simultaneous rotation, scaling, and inversion of SPD matrices. The invariance under matrix inversion is particularly desirable in analysis of DTI data, where both large and small diffusions are unlikely [3]. While these invariance properties are also found in other frameworks [21, 24, 17, 11, 3], the proposed distance is directly interpretable in terms of the relative scaling of eigenvalues and rotation angle between eigenvector frames of two SPD matrices.

For $\mathrm{Sym}^+(3)$, other authors [9, 32] have proposed dissimilarity-measures and interpolation schemes based on the same general idea as ours, i.e., separating the scaling and rotation of SPD matrices. Their deformations of SPD matrices are similar to ours in many cases, thus enjoying similar interpretability. But while [9, 32] mainly focused on the $p = 3$ case, our work is more flexible by allowing *unordered* and *equal* eigenvalues. We discuss the importance of this later in section 3.

The proposed geometric framework for analysis of SPD matrices is viewed as an important first step for developing statistical tools for SPD matrix data that will inherit the interpretability and the advantageous regular behavior of scaling-rotation curves. The development of tools similar to those already existing for other geometric frameworks, such as bi- or trilinear interpolations [3], weighted geometric means and spatial smoothing [21, 10, 7], principal geodesic analysis [11], and regression and statistical testing [34, 28, 27, 33], will also be needed in the new framework, but we do not address them here. The proposed framework also has potential future applications beyond diffusion tensor study, such as in high-dimensional factor models [12] and classification among SPD matrices [15, 30]. Algorithms allowing fast computation or approximation of the proposed distance may be needed, but we will leave this as a subject of future work. The current paper focuses only on analyzing minimal scaling-rotation curves and the distance defined by them.

The main advantage of the new geometric framework for SPD matrices is that minimal scaling-rotation curves interpolate eigenvalues at constant logarithmic rate,

and eigenvectors at constant angular rate, with a minimal amount of scaling and rotation. These are desirable characteristics in fiber-tracking in DTI [5]. Moreover, scaling-rotation curves exhibit regular evolution of determinant and, in typical cases, of fractional anisotropy and mean diffusivity. Linear interpolation of two SPD matrices by the usual vector operation is known to have a *swelling* effect: the determinants of interpolated SPD matrices are larger than those of the two ends. This is physically unrealistic in DTI [3]. The Riemannian frameworks in [21, 24, 3] do not suffer from the swelling effect, which was in part the rationale for favoring the more sophisticated geometry. However, all of these exhibit a *fattening* effect: interpolated SPD matrices are more isotropic than the two ends [8]. The Riemannian frameworks also produce an unpleasant *shrinking* effect: the traces of interpolated SPD matrices are smaller than those of the two ends [5]. The scaling-rotation framework, on the other hand, does not suffer from the fattening effect and produces a smaller shrinking effect with no shrinking at all in the case of pure rotations.

The rest of the paper is organized as follows. Scaling-rotation curves are formally defined in section 2. Section 3 is devoted to precisely characterizing minimal scaling-rotation curves between two SPD matrices and the distance obtained accordingly. The cylindrical representation of $\mathrm{Sym}^+(p)$ is introduced to handle the nonuniqueness of the eigen-decomposition and repeated eigenvalue cases. Section 4 provides details for the computation of the distance and curves for the special but most commonly useful cases of 2×2 and 3×3 SPD matrices. In section 5, we highlight the advantageous regular evolution of the scaling-rotation interpolations of SPD matrices. Technical details, including proofs of theorems, are contained in Appendix B.

2. Scaling-rotation curves in $\mathbf{Sym}^+(p)$. An SPD matrix $M \in \mathrm{Sym}^+(p)$ can be identified with an ellipsoid in $\mathbb{R}^p$ (ellipse if $p = 2$). In particular, the surface coordinates $x \in \mathbb{R}^p$ of the ellipsoid corresponding to M satisfy $x'M^{-1}x = 1$. The semiprincipal axes of the ellipsoid are given by eigenvector and eigenvalue pairs of M. Figure 1 illustrates some SPD matrices in $\mathrm{Sym}^+(3)$ as ellipsoids in $\mathbb{R}^3$. Any deformation of the SPD matrix X to another SPD matrix can be achieved by the combination of two operations:

1. individual scaling of the eigenvalues, or stretching (shrinking) the ellipsoid along principal axes;
2. rotation of the eigenvectors, or rotation of the ellipsoid.

Denote an eigen-decomposition of X by $X = UDU'$, where the columns of $U \in \mathrm{SO}(p)$ consist of orthogonal eigenvectors of X, and $D \in \mathrm{Diag}^+(p)$ is the diagonal matrix of positive eigenvalues that need not be ordered. Here, $\mathrm{SO}(p)$ denotes the set of $p \times p$ real rotation matrices. To parameterize scaling and rotation, the matrix exponential and logarithm, defined in Appendix A, are used. A continuous scaling of the eigenvalues in D at a constant proportionality rate can be described by a curve $D(t) = \exp(Lt)D$ in $\mathrm{Diag}^+(p)$ for some $L = \mathrm{diag}(l_1, \ldots, l_p) \in \mathrm{Diag}(p)$, $t \in \mathbb{R}$, where $\mathrm{Diag}(p)$ is the set of all $p \times p$ real diagonal matrices. Since $\frac{d}{dt}D(t) = LD(t)$, we call L the *scaling velocity*. Each element l_i of L provides the scaling factor for the ith coordinate d_i of D. A rotation of the eigenvectors in the ambient space at a constant "angular rate" is described by a curve $U(t) = \exp(At)U$ in $\mathrm{SO}(p)$, where $A \in \mathfrak{so}(p)$, the set of antisymmetric matrices (the Lie algebra of $\mathrm{SO}(p)$). Since $\frac{d}{dt}U(t) = AU(t)$, we call A the *angular velocity*. Incorporating the scaling and rotation together results in the general scaling-rotation curve (introduced in [25]),

$$\chi(t) = \chi(t; U, D, A, L) = \exp(At)UD\exp(Lt)U'\exp(A't) \in \mathrm{Sym}^+(p), \quad t \in \mathbb{R}. \tag{2.1}$$

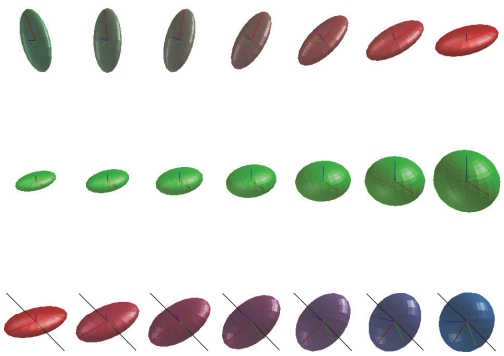

FIG. 1. *Scaling-rotation curves in* $\mathrm{Sym}^+(3)$*: (top) pure rotation with rotation axis normal to the screen, (middle) individual scaling along principal axes without any rotation, and (bottom) simultaneous scaling and rotation. The rotation axis is shown as a black line segment. The ellipsoids are colored by the direction of principal axes, to help visualize the effect of rotation.*

The scaling-rotation curve characterizes deformations of $X = \chi(0) \in \mathrm{Sym}^+(p)$ so that the ellipsoid corresponding to X is smoothly rotated, and each principal axis stretched and shrunk, as a function of t. For $p = 2, 3$, the matrix A gives the axis and angle of rotation (cf. Appendix A). Figure 1 illustrates discretized trajectories of scaling-rotation curves in $\mathrm{Sym}^+(3)$, visualized by the corresponding ellipsoids. These curves in general do not coincide with straight lines or geodesics in other geometric frameworks, such as those in [31, 21, 24, 17, 11, 3, 10, 9, 32]. In section 3, we introduce a Riemannian metric which reproduces these scaling-rotation curves as images of geodesics.

Given two points $X, Y \in \mathrm{Sym}^+(p)$, we will define the distance between them as the length of a scaling-rotation curve $\chi(t)$ that joins X and Y. Thus it is of interest to identify the parameters of the curve $\chi(t)$ that starts at $X = \chi(0)$ and meets $Y = \chi(1)$ at $t = 1$. From eigen-decompositions of X and Y, $X = UDU'$, $Y = V\Lambda V'$, we could equate $\chi(1)$ and $V\Lambda V'$, and naively solve for eigenvector matrix and eigenvalue matrix separately, leading to $A = \log(VU') \in \mathfrak{so}(p)$, $L = \log(D^{-1}\Lambda) \in \mathrm{Diag}(p)$. This solution is generally correct if the eigen-decompositions of X and Y are chosen carefully (see Theorem 3.14). The difficulty is that there are many other scaling-rotation curves that also join X and Y, due to the nonuniqueness of eigen-decomposition. Thus we consider a *minimal* scaling-rotation curve among all such curves.

3. Minimal scaling-rotation curves in $\mathrm{Sym}^+(p)$.

3.1. Decomposition of SPD matrices into scaling and rotation components. An SPD matrix X can be eigen-decomposed into a matrix of eigenvectors $U \in \mathrm{SO}(p)$ and a diagonal matrix $D \in \mathrm{Diag}^+(p)$ of eigenvalues. In general, there are many pairs (U, D) such that $X = UDU'$. Denote the set of all pairs (U, D) by

$$(\mathrm{SO} \times \mathrm{Diag}^+)(p) = \mathrm{SO}(p) \times \mathrm{Diag}^+(p).$$

We use the following notation.

DEFINITION 3.1. *For all pairs* $(U, D) \in (\mathrm{SO} \times \mathrm{Diag}^+)(p)$ *such that* $X = UDU'$*,*

(i) *an eigen-decomposition* (U, D) *of* X *is called an (unobservable)* version *of* X *in* $(\mathrm{SO} \times \mathrm{Diag}^+)(p)$*;*

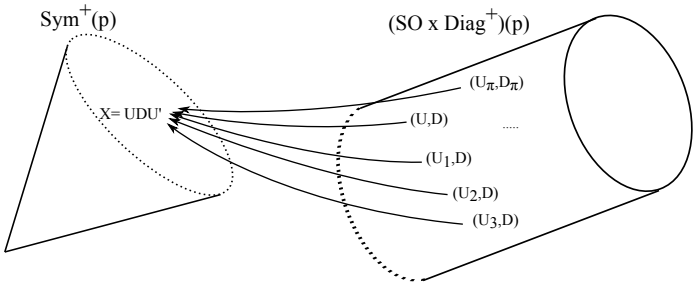

FIG. 2. *An SPD matrix X and its versions in* $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$. *The eigen-composition of* (U,D) *is depicted as a many-to-one mapping from* $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$ *to* $\mathrm{Sym}^+(p)$.

(ii) *X is the* eigen-composition *of (U,D), defined by a mapping* $c:(\mathrm{SO}\times\mathrm{Diag}^+)(p)\to\mathrm{Sym}^+(p)$, $c(U,D)=UDU'=X$.

The many-to-one mapping c from $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$ to $\mathrm{Sym}^+(p)$ is surjective. (The symbol c stands for *composition*.) Figure 2 illustrates the relationship between an SPD matrix and its many versions (eigen-decompositions). While $\mathrm{Sym}^+(p)$ is an open cone, the set $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$ can be understood as the boundary of a generalized cylinder; i.e., $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$ forms a shape of cylinder whose cross-section is "spherical" $(\mathrm{SO}(p))$ and the centers of the cross-section are on the positive orthant of $\mathbb{R}^p$, i.e., $\mathrm{Diag}^+(p)$. The set $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$ is a complete Riemannian manifold, as described below in section 3.2.

Note that considering $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$ as the set of all possible eigen-decompositions is an important relaxation of the usual ordered eigenvalue assumption. We will see in the subsequent sections that this is necessary to describe the desired family of deformations. As an example, the scaling-rotation curve depicted in the middle row of Figure 1 is made possible by allowing *unordered* eigenvalues. Moreover, our manifold $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$ has no boundaries, which not only allows us to handle *equal* eigenvalues but also makes the applied Riemannian geometry simple.

We first discuss which elements of $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$ are the versions of any given SPD matrix X.

DEFINITION 3.2. *Let S_p denote the symmetric group, i.e., the group of permutations of the set $\{1,\dots,p\}$, for $p\geq 2$. A permutation $\pi\in S_p$ is a bijection $\pi:\{1,\dots,p\}\to\{1,\dots,p\}$. Let $\boldsymbol{\sigma}_p=\{(\epsilon_1,\dots,\epsilon_p)\in\mathbb{R}^p:\epsilon_i\in\{\pm1\},1\leq i\leq p\}$ and $\boldsymbol{\sigma}_p^+=\{(\epsilon_1,\dots,\epsilon_p)\in\boldsymbol{\sigma}_p:\prod_{i=1}^p\epsilon_i=1\}$.*

(i) *For a permutation $\pi\in S_p$, its* permutation matrix *is the $p\times p$ matrix P_π^0 whose entries are all 0 except that in column i the entry $\pi(i)$ equals 1. Moreover, define $P_\pi=P_\pi^0$ if $\det(P_\pi^0)=1$, and $P_\pi=\left[\begin{smallmatrix}-1 & \mathbf{0}'\\ \mathbf{0} & I_{p-1}\end{smallmatrix}\right]P_\pi^0$ if $\det(P_\pi^0)=-1$.*

(ii) *For $\sigma=(\epsilon_1,\dots,\epsilon_p)\in\boldsymbol{\sigma}_p$, its associated* sign-change matrix *is the $p\times p$ diagonal matrix I_σ whose ith diagonal element is ϵ_i. If $\sigma\in\boldsymbol{\sigma}_p^+$, we call I_σ an* even sign-change matrix.

(iii) *For any $D\in\mathrm{Diag}(p)$, the* stabilizer subgroup *of D is $G_D=\{R\in\mathrm{SO}(p):RDR'=D\}$.*

For any $\sigma\in\boldsymbol{\sigma}_p^+$, $\pi\in S_p$, $P_\pi,I_\sigma\in\mathrm{SO}(p)$. The number of different permutations (or sign-changes) is $p!$ (or 2^{p-1}, respectively). These two types of matrices provide operations for permutation and sign-changes in eigenvalue decomposition. In particular, for $U\in\mathrm{SO}(p)$, a column-permuted U, by a permutation $\pi\in S_p$, is $UP_\pi'\in\mathrm{SO}(p)$,

and a sign-changed U, by $\sigma \in \boldsymbol{\sigma}_p^+$, is $UI_\sigma \in \mathrm{SO}(p)$. For $D = \mathrm{diag}(d_1, \ldots, d_p)$, define $\pi \cdot D = \mathrm{diag}(d_{\pi^{-1}(1)}, \ldots, d_{\pi^{-1}(p)}) \in \mathrm{Diag}(p)$ as a diagonal matrix whose elements are permuted by $\pi \in S_p$. $D_\pi := P_\pi D P_\pi'$ is exactly the diagonal matrix $\pi \cdot D$. The same is true if P_π is replaced by $I_\sigma P_\pi$ for any $\sigma \in \boldsymbol{\sigma}_p$. Finally, for any $\sigma \in \boldsymbol{\sigma}_p^+$, $\pi \in S_p$, there exists $\sigma^0 \in \boldsymbol{\sigma}_p$ such that $I_{\sigma^0} P_\pi^0 = I_\sigma P_\pi$.

THEOREM 3.3. *Each version of $X = UDU'$ is of the form $(U^*, D^*) = (URP_\pi', D_\pi)$ for $R \in G_D$, $\pi \in S_p$, and $D_\pi = P_\pi D P_\pi'$. Moreover, if the eigenvalues of X are all distinct, every $R \in G_D$ is an even sign-change matrix I_σ, $\sigma \in \boldsymbol{\sigma}_p^+$.*

Remark 3.4. If the eigenvalues of X are all distinct, there are exactly $p!2^{p-1}$ eigen-decompositions of X. In such a case, all versions of X can be explicitly obtained by application of permutations and sign-changes to any version (U, D) of X.

Remark 3.5. If the eigenvalues of X are not all distinct, there are infinitely many eigen-decompositions of X due to the arbitrary rotation R of eigenvectors. The stabilizer group of D, G_D, to which R belongs in Theorem 3.3, does not depend on particular eigenvalues but only on which eigenvalues are equal. More precisely, for $D = \mathrm{diag}(d_1, \ldots, d_p) \in \mathrm{Diag}^+(p)$, let $\mathcal{J}_D$ be the partition of coordinate indices $\{1, \ldots, p\}$ determined by D, i.e., for which i and j are in the same block if and only if $d_i = d_j$. A block can consist of nonconsecutive numbers. For a partition $\mathcal{J} = \{J_1, \ldots, J_r\}$ with r blocks, let $\{W_1, \ldots, W_r\} = \{\mathbb{R}^{J_1}, \ldots, \mathbb{R}^{J_r}\}$ denote the corresponding subspaces of $\mathbb{R}^p$; $x \in \mathbb{R}^{J_i}$ if and only if the jth coordinate of x is 0 for all $j \notin J_i$. The stabilizer G_D depends only on the partition $\mathcal{J}_D$. Define $G_\mathcal{J} \subset \mathrm{SO}(p)$ by

$$G_\mathcal{J} = \{R \in \mathrm{SO}(p) : RW_i = W_i, 1 \le i \le r\}. \tag{3.1}$$

Then $G_D = G_{\mathcal{J}_D}$. As an illustration, let $D = \mathrm{diag}(1, 1, 2)$. Then $\mathcal{J}_D = \{\{1, 2\}, \{3\}\}$. An example of $R \in G_D$ is a 3×3 block-diagonal matrix where the first 2×2 block is any $R_1 \in \mathrm{SO}(2)$ and the last diagonal element is $r_2 = 1$. Intuitively, RDR' with this choice of R behaves as if the first 2×2 block of D, D_1, is arbitrarily rotated. Since $D_1 = I_2$, rotation makes no difference. Another example is given by setting $R_1 \in \mathrm{O}(2)$ with $\det(R_1) = -1$ and $r_2 = -1$.

3.2. A Riemannian framework for scaling and rotation of SPD matrices. The set of rotation matrices $\mathrm{SO}(p)$ is a $p(p-1)/2$-dimensional smooth Riemannian manifold equipped with the usual Riemannian inner product for the tangent space [13, Chap. 18]. The set of positive diagonal matrices $\mathrm{Diag}^+(p)$ is also a p-dimensional smooth Riemannian manifold. The set $(\mathrm{SO} \times \mathrm{Diag}^+)(p)$, being a direct product of two smooth and complete manifolds, is a complete Riemannian manifold [29, 1]. We state some geometric facts necessary to our discussion.

LEMMA 3.6.

(i) $(\mathrm{SO} \times \mathrm{Diag}^+)(p)$ *is a differentiable manifold of dimension* $p + p(p-1)/2$.

(ii) $(\mathrm{SO} \times \mathrm{Diag}^+)(p)$ *is the image of* $\mathfrak{so}(p) \times \mathrm{Diag}(p)$ *under the exponential map* $\mathrm{Exp}((A, L)) = (\exp(A), \exp(L))$, $(A, L) \in \mathfrak{so}(p) \times \mathrm{Diag}(p)$.

(iii) *The tangent space* $\tau(I, I)$ *to* $(\mathrm{SO} \times \mathrm{Diag}^+)(p)$ *at the identity* (I, I) *can be naturally identified as a copy of* $\mathfrak{so}(p) \times \mathrm{Diag}(p)$.

(iv) *The tangent space* $\tau(U, D)$ *to* $(\mathrm{SO} \times \mathrm{Diag}^+)(p)$ *at an arbitrary point* (U, D) *can be naturally identified as the set* $\tau(U, D) = \{(AU, LD) : A \in \mathfrak{so}(p), L \in \mathrm{Diag}(p)\}$.

Our choice of Riemannian inner product at (U, D) for the two tangent vectors (A_1U, L_1D) and (A_2U, L_2D) is

$$\begin{aligned} \langle (A_1U, L_1D), (A_2U, L_2D) \rangle_{(U,D)} &= \frac{k}{2} \langle U'A_1U, U'A_2U \rangle + \langle D^{-1}L_1D, D^{-1}L_2D \rangle \\ &= \frac{k}{2} \mathrm{trace}(A_1A_2') + \mathrm{trace}(L_1L_2), \quad k > 0, \end{aligned} \tag{3.2}$$

where $\langle X, Y\rangle$ for $X, Y \in \mathrm{GL}(p)$ denotes the Frobenius inner product $\langle X, Y\rangle = \mathrm{trace}(XY')$. Collard et al. [9] used a structure similar to (3.2), with the scaling factor k being a function of D, to motivate their distance function. We use $k = 1$ for all of our illustrations in this paper. The practical effect of using different values of k is discussed in section 5.2, and examples are given in the supplementary material, which is linked from the main article webpage. For any fixed k, we show that this choice of Riemannian inner product leads to interpretable distances with invariance properties (cf. Proposition 3.7 and Theorem 3.8).

The exponential map from a tangent space $\tau(U, D)$ to $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$ is $\mathrm{Exp}_{(U,D)} : \tau(U, D) \to (\mathrm{SO} \times \mathrm{Diag}^+)(p)$, where

$$\mathrm{Exp}_{(U,D)}((AU, LD)) = (U \exp(U'AU), D \exp(D^{-1}LD)) = (\exp(A)U, \exp(L)D).$$

The inverse of the exponential map is $\mathrm{Log}_{(U,D)} : (\mathrm{SO} \times \mathrm{Diag}^+)(p) \to \tau(U, D)$, where

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

A geodesic in $(\mathrm{SO} \times \mathrm{Diag}^+)(p)$ starting at (U, D) with initial direction $(AU, LD) \in \tau(U, D)$ is parameterized as

$$\gamma(t) = \gamma(t; U, D, A, L) = \mathrm{Exp}_{(U,D)}((AUt, LDt)). \tag{3.3}$$

The inner product (3.2) provides the geodesic distance function on $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$. Specifically, the squared geodesic distance from (U, D) to (V, Λ) is

$$\begin{aligned} d^2\left((U, D), (V, \Lambda)\right) &= \langle (AU, LD), (AU, LD)\rangle_{(U,D)} \\ &= k d_{\mathrm{SO}(p)}(U, V)^2 + d_{\mathcal{D}}(D, \Lambda)^2, \quad k > 0, \end{aligned} \tag{3.4}$$

where $A = \log(VU')$, $L = \log(\Lambda D^{-1})$, $d_{\mathrm{SO}(p)}(U_1, U_2)^2 = \frac{1}{2}\left\|\log(U_2U_1')\right\|_F^2$, $d_{\mathcal{D}}(D_1, D_2)^2 = \left\|\log(D_2D_1^{-1})\right\|_F^2$, and $\|\ \|_F$ is the Frobenius norm.

The geodesic distance (3.4) is a metric, well-defined for any (U, D) and $(V, \Lambda) \in (\mathrm{SO} \times \mathrm{Diag}^+)(p)$, and is the length of the minimal geodesic curve $\gamma(t)$ that joins the two points. Note that for any two points (U, D) and (V, Λ), there are infinitely many geodesics that connect the two points, just like there are many ways of wrapping a cylinder with a string. There is, however, a unique minimal-length geodesic curve that connects (U, D) and (V, Λ) if VU' is not an involution [20]. (A rotation matrix R is an *involution* if $R \neq I$ and $R^2 = I$.) For $p = 2, 3$, R is an involution if it consists of a rotation through angle π, in which case there are exactly two shortest-length geodesic curves. If VU' is an involution, then V and U are said to be antipodal in $\mathrm{SO}(p)$, and the matrix logarithm of VU' is not unique (there is no principal logarithm), but as discussed in Appendix A, $\log(VU')$ means any solution A of $\exp(A) = VU'$ whose Frobenius norm is the smallest among all such A.

PROPOSITION 3.7. *The geodesic distance* (3.4) *on* $(\mathrm{SO} \times \mathrm{Diag}^+)(p)$ *is invariant under simultaneous left and right multiplication by orthogonal matrices, permutations, and scaling: For any* $R_1, R_2 \in O(p)$, $\pi \in S_p$, *and* $S \in \mathrm{Diag}^+(p)$, *and for any* $(U, D), (V, \Lambda) \in (\mathrm{SO} \times \mathrm{Diag}^+)(p)$, *we have* $\big((U, D), (V, \Lambda)\big) = d\big((R_1UR_2, SD_\pi), (R_1VR_2, S\Lambda_\pi)\big)$.

3.3. Scaling-rotation curves as images of geodesics. We can give a precise characterization of scaling-rotation curves using the Riemannian manifold $(\mathrm{SO} \times$

$\mathrm{Diag}^+)(p)$. In particular, any geodesic in $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$ determines a scaling-rotation curve in $\mathrm{Sym}^+(p)$. The geodesic (3.3) gives rise to the scaling-rotation curve $\chi(t)=\chi(t;U,D,A,L)\in\mathrm{Sym}^+(p)$ (2.1) by the eigen-composition $c\circ\gamma=\chi$. On the other hand, a scaling-rotation curve χ corresponds to many geodesics in $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$.

To characterize the family of geodesics corresponding to a single curve $\chi(t)$, the following notation is used. For a partition $\mathcal{J}$ of the set $\{1,\dots,p\}$, $G_{\mathcal{J}}$ denotes the Lie subgroup of $\mathrm{SO}(p)$ defined in (3.1). Let $\mathfrak{g}_{\mathcal{J}}$ denote the Lie algebra of $G_{\mathcal{J}}$. Then

$$\mathfrak{g}_{\mathcal{J}}=\{A\in\mathfrak{so}(p):A_{ij}=0\text{ for }i\not\sim j\}\subset\mathfrak{so}(p),$$

where $i\not\sim j$ if i and j are in different blocks of $\mathcal{J}$. For $D\in\mathrm{Diag}(p)$, recall from Remark 3.5 that $\mathcal{J}_D$ is the partition determined by eigenvalues of D, $G_D=G_{\mathcal{J}_D}$, and define $\mathfrak{g}_D=\mathfrak{g}_{\mathcal{J}_D}$. For $D,L\in\mathrm{Diag}(p)$, let $\mathcal{J}_{D,L}$ be the common refinement of $\mathcal{J}_D$ and $\mathcal{J}_L$ so that i and j are in the same block of $\mathcal{J}_{D,L}$ if and only if $d_i=d_j$ and $l_i=l_j$. Define $G_{D,L}=G_{\mathcal{J}_{D,L}}=G_D\cap G_L$, and let $\mathfrak{g}_{D,L}=\mathfrak{g}_{\mathcal{J}_{D,L}}=\mathfrak{g}_D\cap\mathfrak{g}_L$ be the Lie algebra of $G_{D,L}$. Finally, for $B\in\mathfrak{so}(p)$, let $\mathrm{ad}_B:\mathfrak{so}(p)\to\mathfrak{so}(p)$ be the linear map defined by $\mathrm{ad}_B(C)=[B,C]=BC-CB$.

THEOREM 3.8. *Let (U,D,A,L) be the parameters of a scaling-rotation curve $\chi(t)$ in $\mathrm{Sym}^+(p)$. Let I be a positive-length interval containing 0. Then a geodesic $\gamma:I\to(\mathrm{SO}\times\mathrm{Diag}^+)(p)$ is identified with χ,* i.e., *$\chi(t)=c(\gamma(t))$, for all $t\in I$, if and only if $\gamma(t)=\gamma(t;URP'_\pi,D_\pi,B,L_\pi)$ for some $\pi\in S_p$, $R\in G_{D,L}$, and $B\in\mathfrak{so}(p)$ satisfying both* (i) *$\tilde{B}-\tilde{A}\in\mathfrak{g}_{D,L}$, where $\tilde{B}=U'BU$ and $\tilde{A}=U'AU$, and* (ii) *$(ad_{\tilde{B}})^j(\tilde{A})\in\mathfrak{g}_{D,L}$ for all $j\ge1$.*

Note that the conjugation $\tilde{A}=U'AU$ expresses the infinitesimal rotation parameter A in the coordinate system determined by U. If A in Theorem 3.8 is such that $\tilde{A}\in\mathfrak{g}_{D,L}$, then conditions (i) and (ii) are equivalent to $\tilde{B}\in\mathfrak{g}_{D,L}$. If $p=2$ or 3 and $\tilde{A}\notin\mathfrak{g}_{D,L}$, then the condition is $\tilde{B}=\tilde{A}$.

It is worth emphasizing a special case where there are only finitely many geodesics corresponding to a scaling-rotation curve $\chi(t)$.

COROLLARY 3.9. *Suppose, for some t, the eigenvalues of $\chi(t)=\chi(t;U,D,A,L)$ are distinct. Then χ corresponds to only finitely many $(p!2^{p-1})$ geodesics $\gamma(t)=\gamma(t;UI_\sigma P'_\pi,D_\pi,A,L_\pi)$, where $\pi\in S_p$ and $\sigma\in\boldsymbol{\sigma}_p^+$.*

3.4. Scaling-rotation distance between SPD matrices. In $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$, consider the set of all elements whose eigen-composition is X:

$$\mathcal{E}_X=\{(U,D)\in(\mathrm{SO}\times\mathrm{Diag}^+)(p):X=UDU'\}.$$

Since the eigen-composition is a surjective mapping, the collection of these sets $\mathcal{E}_X$ partitions the manifold $(\mathrm{SO}\times\mathrm{Diag}^+)(p)$. The set $\mathcal{E}_X=c^{-1}(X)$ is called the fiber over X. Theorem 3.3 above characterizes all members of $\mathcal{E}_X$ for any X.

It is natural to define a distance between X and $Y\in\mathrm{Sym}^+(p)$ to be the length of the shortest geodesic connecting $\mathcal{E}_X$ and $\mathcal{E}_Y\subset(\mathrm{SO}\times\mathrm{Diag}^+)(p)$.

DEFINITION 3.10. *For $X,Y\in\mathrm{Sym}^+(p)$, the scaling-rotation distance is defined as*

$$d_{\mathcal{SR}}(X,Y):=\inf_{\substack{(U,D)\in\mathcal{E}_X,\\(V,\Lambda)\in\mathcal{E}_Y}}d((U,D),(V,\Lambda)),\tag{3.5}$$

where $d(\cdot,\cdot)$ is the geodesic distance function (3.4).

The geodesic distance $d((U,D),(V,\Lambda))$ measures the length of the shortest geodesic segment connecting (U,D) and (V,Λ). Any geodesic, mapped to $\mathrm{Sym}^+(p)$ by the eigen-composition, is a scaling-rotation curve connecting $X = UDU'$ and $Y = V\Lambda V'$. In this sense, the scaling-rotation distance $d_{\mathcal{SR}}$ measures the minimum amount of smooth deformation from X to Y (or vice versa) only by the rotation of eigenvectors and individual scaling of eigenvalues.

Note that $d_{\mathcal{SR}}$ on $\mathrm{Sym}^+(p)$ is well-defined and the infimum is actually achieved, as both $\mathcal{E}_X$ and $\mathcal{E}_Y$ are nonempty and compact. It has desirable invariance properties and is a semimetric on $\mathrm{Sym}^+(p)$.

THEOREM 3.11. *For any $X, Y \in \mathrm{Sym}^+(p)$, the scaling-rotation distance $d_{\mathcal{SR}}$ is*

(i) *invariant under matrix inversion, i.e., $d_{\mathcal{SR}}(X,Y) = d_{\mathcal{SR}}(X^{-1},Y^{-1})$;*

(ii) *invariant under simultaneous uniform scaling and conjugation by a rotation matrix, i.e., $d_{\mathcal{SR}}(X,Y) = d_{\mathcal{SR}}(sRXR', sRYR')$ for any $s > 0$, $R \in \mathrm{SO}(p)$;*

(iii) *a semimetric on $\mathrm{Sym}^+(p)$; i.e., $d_{\mathcal{SR}}(X,Y) \geq 0$, $d_{\mathcal{SR}}(X,Y) = 0$ if and only if $X = Y$, and $d_{\mathcal{SR}}(X,Y) = d_{\mathcal{SR}}(Y,X)$.*

Although $d_{\mathcal{SR}}$ is not a metric on the entire set $\mathrm{Sym}^+(p)$, it is a metric on an important subset of $\mathrm{Sym}^+(p)$.

THEOREM 3.12. *$d_{\mathcal{SR}}$ is a metric on the set of SPD matrices whose eigenvalues are all distinct.*

3.5. Minimal scaling-rotation curves in $\mathrm{Sym}^+(p)$. To evaluate the scaling-rotation distance (3.5), it is necessary to find a shortest-length geodesic in $(\mathrm{SO} \times \mathrm{Diag}^+)(p)$ between the fibers $\mathcal{E}_X$ and $\mathcal{E}_Y$. There are multiple geodesics connecting two fibers, because each fiber contains at least $p!2^{p-1}$ elements (Theorem 3.3), as depicted in Figure 3. We think of fibers $\mathcal{E}_X$ arranged vertically in $(\mathrm{SO} \times \mathrm{Diag}^+)(p)$ with the mapping c (eigen-composition) as downward projection. It is clear that there exists a geodesic that joins the two fibers with the minimal distance. We call such a geodesic a *minimal geodesic* for the two fibers $\mathcal{E}_X$ and $\mathcal{E}_Y$. A necessary, but generally not sufficient, condition for a geodesic to be minimal for $\mathcal{E}_X$ and $\mathcal{E}_Y$ is that it is perpendicular to $\mathcal{E}_X$ and $\mathcal{E}_Y$ at its endpoints. A pair $((U,D),(V,\Lambda)) \in \mathcal{E}_X \times \mathcal{E}_Y$ is called a *minimal pair* if (U,D) and (V,Λ) are connected by a minimal geodesic. The distance $d_{\mathcal{SR}}(X,Y)$ is the length of any minimal geodesic segment connecting the fibers $\mathcal{E}_X$ and $\mathcal{E}_Y$.

DEFINITION 3.13. *Let $X, Y \in \mathrm{Sym}^+(p)$. A scaling-rotation curve $\chi : [0,1] \to \mathrm{Sym}^+(p)$, as defined in* (2.1)*, with $\chi(0) = X$ and $\chi(1) = Y$, is called* minimal *if $\chi = c \circ \gamma$ for some minimal geodesic segment γ connecting $\mathcal{E}_X$ and $\mathcal{E}_Y$.*

THEOREM 3.14. *Let $X, Y \in \mathrm{Sym}^+(p)$. Let $((U,D),(V,\Lambda))$ be a minimal pair for X and Y, and let $A = \log(VU')$, $L = \log(D^{-1}\Lambda)$. Then the scaling-rotation curve $\chi(t;U,D,A,L)$, $0 \leq t \leq 1$, is minimal.*

The above theorem tells us that for any two points $X, Y \in \mathrm{Sym}^+(p)$, a minimal scaling-rotation curve is determined by a minimal pair of $\mathcal{E}_X$ and $\mathcal{E}_Y$. Procedures to evaluate the parameters of the minimal rotation-scaling curve and to compute the scaling-rotation distance are provided for the special cases $p = 2, 3$ in section 4.

The minimal scaling-rotation curve may not be unique. The following theorem gives sufficient conditions for uniqueness.

THEOREM 3.15. *Let $((U,D),(V,\Lambda))$ be a minimal pair for $\mathcal{E}_X$ and $\mathcal{E}_Y$, and let $\chi_o(t) = \chi(t;U,D,\log(VU'),\log(D^{-1}\Lambda))$ be the corresponding minimal scaling-rotation curve.*

(i) *If either all eigenvalues of D are distinct or Λ has only one distinct eigenvalue, and if (V,Λ) is the unique minimizer of $d((U,D),(V_0,\Lambda_0))$ among all*

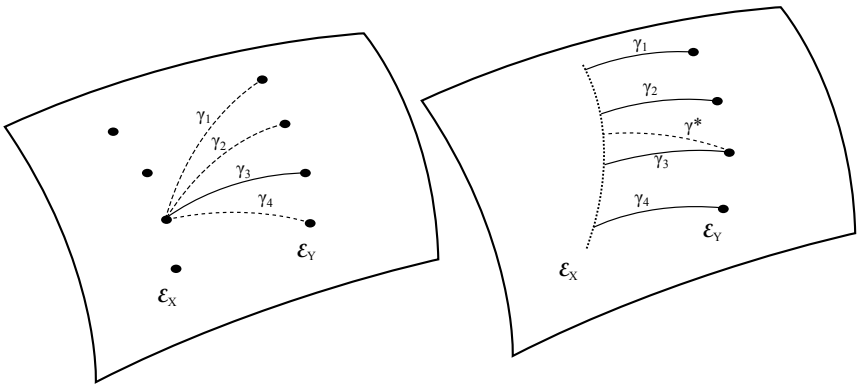

FIG. 3. *(left)* $(\mathrm{SO}\times\mathrm{Diag}^+)(2)$ *is drawn as a curved manifold. In this picture, the four versions of* X *(and of* Y*) are displayed vertically. For a fixed version* (U_3, D_3) *of* X*, there are four geodesics* γ_i *joining* (U_3, D_3) *and the* i*th version of* Y*. A minimal geodesic (*γ_3 *in this figure) has the shortest length among these geodesics. (right) The fiber* $\mathcal{E}_X$ *has infinitely many versions, shown as a vertical dotted curve in* $(\mathrm{SO}\times\mathrm{Diag}^+)(2)$*. There exist multiple minimal geodesics* γ_i *with the shortest length, all of which meet the vertical fiber* $\mathcal{E}_X$ *in a right angle. Here,* $\gamma*$ *is an example of a nonminimal geodesic that does not meet* $\mathcal{E}_X$ *orthogonally.*

$(V_0, \Lambda_0) \in \mathcal{E}_Y$*, then all minimal geodesics between* $\mathcal{E}_X$ *and* $\mathcal{E}_Y$ *are mapped by* c *to the unique* $\chi_o(t)$ *in* $\mathrm{Sym}^+(p)$*.*

(ii) *If there exists* $(V_1, \Lambda_1) \in \mathcal{E}_Y$ *such that* $(V_1, \Lambda_1) \neq (V, \Lambda)$ *and the pair* $((U, D), (V_1, \Lambda_1))$ *is also minimal, then* $\chi_1(t) = \chi(t; U, D, \log(V_1U'), \log(D^{-1}\Lambda_1))$ *is also minimal and* $\chi_1(t) \neq \chi_o(t)$ *for some* t*.*

The following example shows a case with a unique minimal scaling-rotation curve and two cases exhibiting nonuniqueness.

Example. Consider $X = \mathrm{diag}(e, e^{-1})$ and $Y = R_\theta(2X)R'_\theta$, where R_θ is the 2×2 rotation matrix by counterclockwise angle θ.

(i) If $\theta = \pi/3$, then there exists a unique minimal scaling-rotation curve between X, Y. This ideal case is depicted in Figure 4, where among the four scaling-rotation curves, the red curve χ_4 is minimal, as indicated by the length of the curves. In the upper right panel, a version (I, X) of X, depicted as a diamond, and a version of Y are joined by the red minimal geodesic segment.

(ii) Suppose $\theta = \pi/2$. There are two minimal scaling-rotation curves, one by uniform scaling and counterclockwise rotation, the other by the same uniform scaling but by clockwise rotation.

(iii) Let $X = \mathrm{diag}(e^{\epsilon/2}, e^{-\epsilon/2})$ and $Y = R_\theta X R'_\theta$. For $0 \le \theta \le \pi/2$,

$$d_{\mathcal{SR}}(X,Y) = \min\left\{\theta, \sqrt{\left(\frac{\pi}{2}-\theta\right)^2 + 2\epsilon^2}\right\} = \begin{cases} \theta, & \theta \le \frac{\pi}{4} + \frac{2\epsilon^2}{\pi}, \\ \sqrt{(\frac{\pi}{2}-\theta)^2 + 2\epsilon^2} & \text{otherwise.} \end{cases}$$

If the rotation angle is less than 45 degrees or the SPD matrices are highly anisotropic (large ϵ), then the minimal scaling-rotation is a pure rotation (leading to the distance θ). On the other hand, if the matrices are close to being isotropic (eigenvalues ≈ 1), the minimal scaling-rotation curve is given by simultaneous rotation and scaling. An exceptional case arises when $\theta = \frac{\pi}{4} + \frac{2\epsilon^2}{\pi} < \frac{\pi}{2}$, where both curves are of the same length, and there are two minimal scaling-rotation curves.

4. Computing the minimal scaling-rotation curve and scaling-rotation distance. We provide computation procedures for the scaling-rotation distance

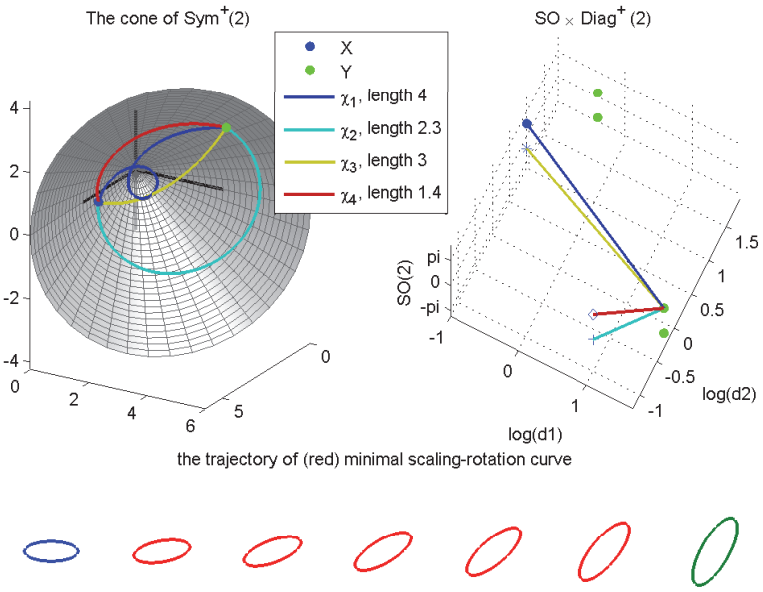

FIG. 4. *Two SPD matrices X (blue) and Y (green) in the cone of* $\mathrm{Sym}^+(2)$ *(top left), and their four versions in a flattened* $(\mathrm{SO}\times\mathrm{Diag}^+(2))$ *(top right). The eigen-composition of each shortest geodesic connecting versions of X and Y is a scaling-rotation curve in* $\mathrm{Sym}^+(2)$*. Different colors represent four different such curves. The red scaling-rotation curve has the shortest geodesic distance in* $(\mathrm{SO}\times\mathrm{Diag}^+)(2)$ *and thus is minimal. Its trajectory is shown as the deformation of ellipses in the bottom panel (from leftmost X to rightmost Y). Color is available only in the online version.*

$d_{\mathcal{SR}}(X,Y)$ for $X,Y\in\mathrm{Sym}^+(2)$ or $\mathrm{Sym}^+(3)$. Theorems 4.1 and 4.3 below provide the minimal pair(s), based on which the exact formulation of the minimal scaling-rotation curve is evaluated in Theorem 3.14 above.

4.1. Scaling-rotation distance for 2×2 SPD matrices. Let (d_1,d_2) be the eigenvalues of X, and let (λ_1,λ_2) be the eigenvalues of Y.

THEOREM 4.1. *Given any 2×2 SPD matrices X and Y, the distance* (3.5) *is computed as follows.*

(i) *If $d_1\neq d_2$ and $\lambda_1\neq\lambda_2$, then there are exactly four versions of X, denoted by (U_i,D_i), $i=1,\dots,4$, and for any version (V,Λ) of Y,*

$$d_{\mathcal{SR}}(X,Y)=\min_{i=1,\dots,4} d((U_i,D_i),(V,\Lambda)). \tag{4.1}$$

These versions are given by the permutation and sign changes.

(ii) *If $d_1=d_2$, then for any version (V,Λ) of Y, $d_{\mathcal{SR}}(X,Y)=d((V,D),(V,\Lambda))$, regardless of whether the eigenvalues of Y are distinct or not.*

Therefore, the minimizer of (U_o,D_o) of (4.1) and (V,Λ) are a minimal pair for case (i); $(V,D),(V,\Lambda)$ are a minimal pair for case (ii).

4.2. Scaling-rotation distance for 3×3 SPD matrices. Let $X,Y\in\mathrm{Sym}^+(3)$. Let (d_1,d_2,d_3) be the eigenvalues of X, and $(\lambda_1,\lambda_2,\lambda_3)$ the eigenvalues of Y, without any given ordering. In order to separately analyze and catalogue all cases of eigenvalue multiplicities in Theorem 4.3 below, we will use the following details for the case where an eigenvalue of X is of multiplicity 2.

For any version (U, D) with $D = \mathrm{diag}(d_1, d_1, d_3)$, $d_1 = d_2$, all other versions of X are of the form $(UR_1P'_\pi, D_\pi)$ for permutation π and rotation matrix $R_1 \in G_D$ (Theorem 3.3). We can take $R_1 = RI_\sigma$ for some $\sigma \in \boldsymbol{\sigma}_p^+$ and for some diagonal rotation matrix R with $+1$ in the lower right-hand corner. For fixed (U, D), (V, Λ), $\sigma \in \boldsymbol{\sigma}_p^+$, and $\pi \in S_p$, one can find a *minimal rotation* $\hat{R}_{\sigma,\pi}$ satisfying

$$d((U\hat{R}_{\sigma,\pi}I_\sigma P'_\pi, D_\pi), (V, \Lambda)) \le d((URI_\sigma P'_\pi, D_\pi), (V, \Lambda))$$

for all such R, as the following lemma states.

LEMMA 4.2. *Let* $\Gamma = I_\sigma P'_\pi V'U = \left[\begin{smallmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \gamma_{22} \end{smallmatrix}\right]$, *where* Γ_{11} *is the first* 2×2 *block of* Γ. *The minimal rotation matrix* $\hat{R}_{\sigma,\pi} = \hat{R}$ *is given by* $\hat{R} = \left[\begin{smallmatrix} E_2E'_1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, *where* $E_1\Lambda_\Gamma E'_2$ *is the "semisingular values" decomposition of* Γ_{11}. *(In semisingular values decomposition, we require that* $E_1, E_2 \in \mathrm{SO}(2)$ *and that the diagonal entries* λ_1 *and* λ_2 *of* Λ_Γ *satisfy* $\lambda_1 \ge |\lambda_2| \ge 0$.*)*

Each choice of σ and π produces a *minimally rotated version* $(\hat{U}_{\sigma,\pi}, D_\pi) = (U\hat{R}_{\sigma,\pi}I_\sigma P'_\pi, D_\pi)$. To provide a minimal pair as needed in Theorem 3.14, a combinatorial problem involving the $3!2^{3-1} = 24$ choices of (σ, π) needs to be solved since the version of X closest to (V, Λ) is found by comparing distances between $(\hat{U}_{\sigma,\pi}, D_\pi)$ and (V, Λ). Fortunately, there are only six such minimally rotated versions corresponding to six choices of (σ, π). In particular, we need only $\pi_1 : (1, 2, 3) \to (1, 2, 3)$, $\pi_2 : (1, 2, 3) \to (3, 1, 2)$, $\pi_3 : (1, 2, 3) \to (1, 3, 2)$, and $\sigma_1 = (1, 1, 1)$, $\sigma_2 = (-1, 1, -1)$, and $\hat{R}_{\sigma_j,\pi_i}$ can be found for each (σ_j, π_i), $i = 1, 2, 3$, $j = 1, 2$. The other pairs of permutations and sign-changes do not need to be considered because each of them will produce one of the six minimally rotated versions, with the same distance from (V, Λ).

THEOREM 4.3. *Given any* 3×3 *SPD matrices* X *and* Y, *the distance* (3.5) *is computed as follows:*

(i) *If the eigenvalues of* X *(and also of* Y*) are all distinct, then there are exactly* 24 *versions of* X, *denoted by* (U_i, D_i), $i = 1, \ldots, 24$, *and for any version* (V, Λ) *of* Y, $d_{\mathcal{SR}}(X, Y) = \min_{i=1,\ldots,24} d((U_i, D_i), (V, \Lambda))$.

(ii) *If* $d_1 = d_2 \ne d_3$ *and* $\{\lambda_1, \lambda_2, \lambda_3\}$ *are distinct, then for any version* (V, Λ) *of* Y *and a version* (U, D) *of* X *satisfying* $D = \mathrm{diag}(d_1, d_1, d_3)$,

$$d_{\mathcal{SR}}(X, Y) = \min_{i=1,2,3,\, j=1,2} d((\hat{U}_{\sigma_j,\pi_i}, D_{\pi_i}), (V, \Lambda)),$$

where $(\hat{U}_{\sigma_j,\pi_i}, D_{\pi_i})$, $i = 1, 2, 3$, $j = 1, 2$, *are the six minimally rotated versions.*

(iii) *If* $d_1 = d_2 \ne d_3$ *and* $\lambda_1 = \lambda_2 \ne \lambda_3$, *then choose* $D = \mathrm{diag}(d_1, d_2, d_3)$ *and* $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$. *For any versions* (U, D), (V, Λ) *of* X *and* Y,

$$d_{\mathcal{SR}}(X, Y) = \min_{i=1,2,3,\, j=1,2} d((UR_{\theta_{ij}}I_{\sigma_i}P'_{\pi_j}, D_{\pi_j}), (VR_{\phi_{ij}}, \Lambda)),$$

where $R_\theta = \exp([a]_\times)$, $a = (0, 0, \theta)'$ *(cf. Appendix* A*), and* (θ_{ij}, ϕ_{ij}) *simultaneously maximizes* $G(\theta, \phi) = \mathrm{trace}(UR_\theta I_{\sigma_i}P'_{\pi_j}R'_\phi V')$.

(iv) *If* $d_1 = d_2 = d_3$, *then for any version* (V, Λ) *of* Y, *we have* $d_{\mathcal{SR}}(X, Y) = d((V, D), (V, \Lambda))$, *regardless of whether the eigenvalues of* Y *are distinct or not.*

The minimizer (θ_{ij}, ϕ_{ij}) of $G(\theta, \phi)$ in Theorem 4.3(iii) is found by a numerical method. Specifically, given the mth iterates $\theta^{(m)}, \phi^{(m)}$, the $(m+1)$st iterate $\theta^{(m+1)}$ is

the solution θ in Lemma 4.2, treating $VR_{\phi^{(m)}}$ as V. We then find $\phi^{(m+1)}$ similarly by using Lemma 4.2, with the role of U and V switched. In our experiments, convergence to the unique maximum was fast and reached by only a few iterations.

5. Scaling-rotation interpolation of SPD matrices. For $X, Y \in \mathrm{Sym}^+(p)$, *a scaling-rotation interpolation from X to Y* is defined as any minimal scaling-rotation curve $f_{SR}(t) := \chi_o(t)$, $t \in [0,1]$, such that $f_{SR}(0) = X$, $f_{SR}(1) = Y$. By definition, every scaling-rotation curve $\chi(t; U, D, A, L)$, and hence every scaling-rotation interpolation, has a log-constant scaling velocity L and constant angular velocity A. The scalar $\mathrm{trace}(L)$ gives the (constant) speed at which log-determinant evolves: $\log(\det \chi(t)) = \log(\det(D)) + \mathrm{trace}(L)t$. Analogously, we view the scalar quantity $\|A\|_F / \sqrt{2}$ as a constant *speed of rotation*, and for all $t \geq 0$ we define the *amount of rotation* applied from time 0 to time t to be $\theta_t := t \|A\|_F / \sqrt{2}$. For a minimal pair $((U, D), (V, \Lambda))$ of X and Y, and the corresponding scaling-rotation interpolation f_{SR}, we have

$$\log(\det f_{SR}(t)) = (1-t)\log(\det(X)) + t\log(\det(Y)), \tag{5.1}$$

and we define the *amount of rotation applied by f_{SR} from X to Y* to be $\theta := \|\log(VU')\|_F / \sqrt{2}$. For $p = 2, 3$, θ is equal to the angle of rotation.

5.1. An application to diffusion tensor computing. This work provides an interpretative geometric framework in analysis of diffusion tensor magnetic resonance images [16], where diffusion tensors are given by 3×3 SPD matrices. Interpolation of tensors is important for fiber tracking, registration, and spatial normalization of diffusion tensor images [5, 8]. The scaling-rotation curve can be understood as a deformation path from one diffusion tensor to another, and is nicely interpreted as scaling of diffusion intensities and rotation of diffusion directions. This advantage in interpretation has not been found in popular geometric frameworks such as [24, 11, 3, 10, 6]. The approaches in [5, 8, 9, 32] also explicitly use rotation of directions, and many scaling-rotation curves are very similar to the deformation paths given in [9, 32]. We defer the discussion on the difference between our framework and those in [9, 32] to section 5.2.

As an example, consider interpolating from $X = \mathrm{diag}(15, 2, 1)$ to Y, whose eigenvalues are $(100, 2, 1)$ and whose principal axes are different from those of X. The first row of Figure 5 presents the corresponding evolution ellipsoids by the scaling-rotation interpolation f_{SR}. This evolution is consistent with human perception when deforming X to Y. As shown in the two bottom left panels of Figure 5, the interpolation exhibits the constant angular rate of rotation and log-constant rate of change of determinant.

By way of comparison, the Euclidean interpolation in row 2 is defined by $f_E(t) = (1-t)X + tY$. The log-Euclidean and affine-invariant Riemannian interpolations in rows 3 and 4 are defined by $f_{LE}(t) = \exp((1-t)\log(X) + t\log(Y))$ and $f_{AI}(t) = X^{\frac{1}{2}} \exp(t \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}})) X^{\frac{1}{2}}$, respectively; see [3]. For these interpolations, we define the rotation angle at time t by the angle of swing from the major axis at time 0 to that at time t. These rotation angles are not, in general, linear in t, as the bottom left panel illustrates. The log-Euclidean f_{LE} and affine-invariant interpolations f_{AI} are log-linear in determinant, and in fact (5.1) holds exactly for f_{LE} and f_{AI}. On the other hand, the Euclidean interpolation is known to suffer from the *swelling effect*: $\det(f_E(t)) > \max(\det(X), \det(Y))$ for some $t \in [0,1]$ [3]. This is shown in the second bottom panel of Figure 5 for the same example. The other interpolations, f_{SR}, f_{LE}, and f_{AI}, do not suffer from the swelling effect.

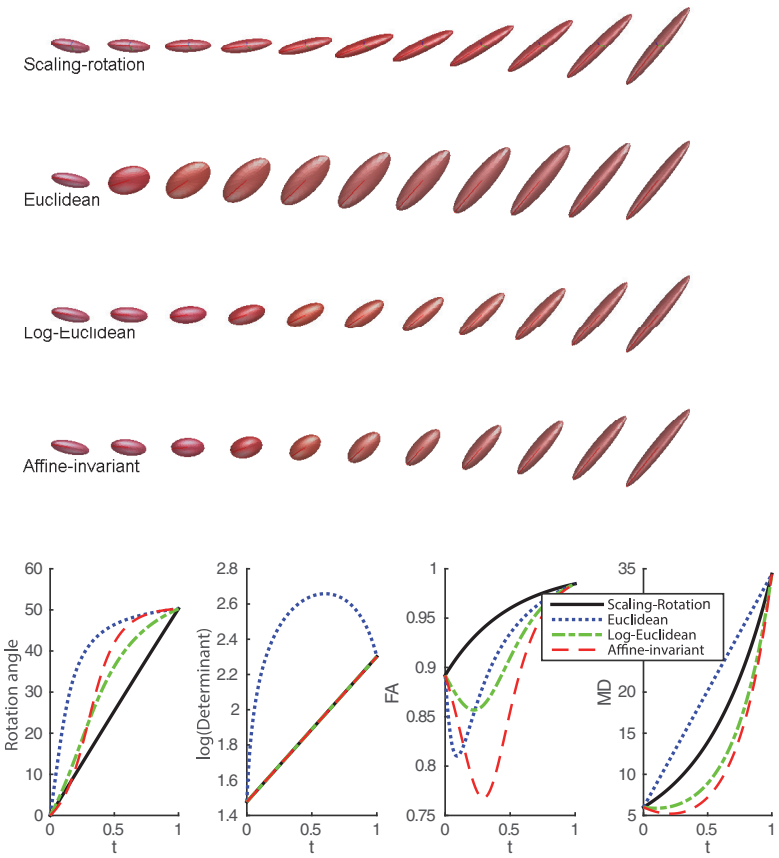

FIG. 5. *(Top) Interpolations of two* 3×3 *SPD matrices. Row* 1*: Scaling-rotation interpolation by the minimal scaling-rotation curve. Row* 2*: (Euclidean) linear interpolation on coefficients. Row* 3*: Log-Euclidean geodesic interpolation. Row* 4*: Affine-invariant Riemannian interpolation. The pointy shape of ellipsoids on both ends is well preserved in the scaling-rotation interpolation. (Bottom) Evolution of rotation angle, determinant, FA, and MD for these four interpolations. Only the scaling-rotation interpolation provides a monotone pattern.*

Minimal scaling-rotation curves not only provide regular evolution of rotation angles and determinant, they also minimize the combined amount of scaling and rotation, as in Definition 3.13. This results in a particularly desirable property: in many examples, the fractional anisotropy (FA) and mean diffusivity (MD) evolve monotonically. FA measures a degree of anisotropy that is zero if all eigenvalues are equal and approaches 1 if one eigenvalue is held constant and the other two approach zero; see [16]. MD is the average of eigenvalues $\mathrm{MD}(X) = \mathrm{trace}(X)/3$.

In the example of Figure 5, $\mathrm{FA}(f_{SR}(t))$ increases monotonically. In contrast, other interpolations of the highly anisotropic X and Y become less anisotropic. This phenomenon may be called a *fattening effect*: interpolated SPD matrices are more isotropic than the two ends [8]. Moreover, log-Euclidean and affine-invariant Riemannian interpolations can suffer from a *shrinking effect*: the MD of interpolated SPD matrices is smaller than the MD of either end [5], as shown in the bottom right panel. In this example, the scaling-rotation interpolation does not suffer from fattening and

shrinking effects. These adverse effects are less severe in f_{SR} than in f_{LE} or f_{AI} in most typical examples, as shown in the supplementary material.

The advantageous regular evolution results from the rotational part of the interpolation. To see this, consider, as in [5], a case where the interpolation by f_{SR} consists only of rotation (a precise example is shown in Figure 1 in the supplementary material). The scaling-rotation interpolation preserves the determinant, FA, and MD, while the other modes of interpolation exhibit irregular behavior in some of the measurements. On the other hand, when $f_{SR}(t)$ is composed of pure scaling, then $f_{SR}(t) = f_{LE}(t) = f_{AI}(t)$ for all t, and there is no guarantee that MD or FA grows monotonically for either curve. The equality of these three curves in this special case is a consequence of the geometric scaling of eigenvalues in the scaling-rotation curve (2.1), which in turn is a consequence of our use of the Riemannian inner product (3.2).

In summary, while the three most popular methods suffer from swelling, fattening, or shrinking effects, the scaling-rotation interpolation provides good regular evolution of all three summary statistics, and solely provides constant angular rate of rotation. More examples illustrating these effects in various scenarios are given in the supplementary material.

5.2. Comparing to other rotation-scaling schemes. Geometric frameworks for 3×3 SPD matrices that decouple rotation from scaling have also been developed in [9, 32]. However, our framework differs in two major ways. First, we allow unordered and equal eigenvalues in any dimension, while in [9, 32] only dimension 3 and only the case of distinct, ordered eigenvalues were considered. In our framework, every scaling-rotation curve corresponds to geodesics in a smooth manifold, which is not possible if eigenvalues are ordered. This leads to a more flexible family of interpolations than those of [9, 32], as we illustrate in the supplementary material.

Other differences lie in the choice of the metric for SO(3) and the weight k in (3.2). While we use geodesic distance and interpolation determined by the standard Riemannian metric on SO(3), in [9] a chordal distance and extrinsic interpolation were used. As a consequence, the interpolation in [9] is close to, but not equal to, a special case of minimal scaling-rotation curves, in particular when k in (3.2) is small. An example illustrating this effect of k is given in section 2 of the supplementary material.

Appendix A. Parameterization of scaling and rotation. The matrix exponential of a square matrix Y is $\exp(Y) = \sum_{j=0}^{\infty} \frac{Y^j}{j!}$. For a square matrix X, if there exists a unique matrix Y of smallest norm such that $X = \exp(Y)$, then we call Y the principal logarithm of X, denoted $\log(X)$.

The exponential map from $\mathrm{Diag}(p)$ to $\mathrm{Diag}^+(p)$, defined by the matrix exponential, is bijective. Moreover, the elementwise exponential and logarithm for the diagonal elements give the matrix exponential and logarithm for $\mathrm{Diag}(p)$ and $\mathrm{Diag}^+(p)$ [13, Chap. 18].

Rotation matrices can be parameterized by antisymmetric matrices since the exponential map from $\mathfrak{so}(p)$ to $\mathrm{SO}(p)$ is onto. Contrary to the $\mathrm{Diag}^+(p)$ case, the matrix exponential is not one-to-one. The principal logarithm is defined on the set $\{R \in \mathrm{SO}(p) : R \text{ is not an involution}\}$, a dense open subset of $\mathrm{SO}(p)$. For completeness, when there exists no principal logarithm of R, we use the notation $\log(R)$ to denote any solution A of $\exp(A) = R$ satisfying that $\|A\|_F$ is the smallest among all such choices of A.

This parameterization gives a physical interpretation of rotations. Specifically, in

the case of $p = 3$, a rotation matrix $R = \exp(A) \in \mathrm{SO}(3)$ can be understood as a linear operator rotating a vector in the real 3-space around an "axis" $a = (a_1, a_2, a_3)'$ by angle $\theta = \|a\|_2$ (in radians), where $A \in \mathfrak{so}(3)$ is the cross-product matrix of a defined by

$$A = [a]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \in \mathfrak{so}(3).$$

Explicit formulas for the matrix exponential and logarithm are given in the following.

LEMMA A.1 (see [20, 13]).

(i) *(Rodrigues' formula) Any* $A \in \mathfrak{so}(3)$ *equals* $[a]_\times$ *for some axis* $a = \theta\tilde{a} \in \mathbb{R}^3$, $\|\tilde{a}\| = 1$. *An explicit formula for the matrix exponential of* A *is* $\exp(A) = I + [\tilde{a}]_\times \sin(\theta) + [\tilde{a}]_\times^2 (1 - \cos(\theta))$.

(ii) *For any* $R \in \mathrm{SO}(3)$, *there exists* $\theta \in [0, \pi]$ *satisfying* $2\cos(\theta) = \mathrm{trace}(R) - 1, \|\log(R)\|_F = \sqrt{2}|\theta|$. *If* $\theta < \pi$, *then* R *has the principal logarithm* $\log(R) = \frac{\theta}{2\sin(\theta)}(R - R')$ *if* $\theta \in (0, \pi)$ *and* 0 *if* $\theta = 0$.

If there exists no principal logarithm of R, i.e., if $\theta = \pi$, then we take $\log(R)$ to be either of the two elements $A \in \mathfrak{so}(3)$ satisfying $\exp(A) = R$ and $\|A\|_F = \sqrt{2}\pi$.

Appendix B. Additional lemmas and proofs.

Proof of Theorem 3.3. For any $(U^*, D^*) \in (\mathrm{SO} \times \mathrm{Diag}^+)(p)$, there exists $V \in \mathrm{SO}(p)$, $L \in \mathrm{Diag}^+(p)$ such that $U^* = UV$ and $D^* = DL$. Therefore, a version of (U, D) can be written as (UV, DL) satisfying $UVDLV'U' = UDU'$, or equivalently,

$$VDLV' = D. \tag{B.1}$$

The set of eigenvalues of $VDLV'$ is $\{d_i l_i : i = 1, \ldots, p\}$, which should be the same as the eigenvalues of D. That is, $d_i l_i = d_j$ for some j. In other words, for some permutation π, $DL = D_\pi$. There are at most $p!$ possible ways to achieve this.

Observe that there exists $R \in \mathrm{SO}(p)$ such that $RP_\pi' = V$ for any V and π. Equation (B.1) is then $RP_\pi' D_\pi P_\pi R' = D$, which becomes $RDR' = D$ since $P_\pi' D_\pi P_\pi = D$.

The last statement of Theorem 3.3 can be seen from noting that R and D commute, so the eigenvector matrix of R is I, with eigenvalues $\{e^{i\theta_j}, e^{-i\theta_j}, 1 : j = 1, \ldots, \lfloor p/2 \rfloor\}$ [14, Cor. 2.5.11]. However, all possible values of θ_j are either 0 or π because R must be a real matrix. Therefore, R is an even sign-change matrix. ∎

Proof of Proposition 3.7. Since S commutes with other diagonal matrices,

$$\begin{aligned} & d^2\left((R_1 U R_2, S D_\pi), (R_1 V R_2, S\Lambda_\pi)\right) \\ &= \frac{k}{2}\left\|\log(R_1 U R_2 R_2' V' R_1')\right\|_F^2 + \mathrm{trace}(\log^2(S\Lambda_\pi D_\pi^{-1} S^{-1})) \\ &= \frac{k}{2}\left\|\log(UV')\right\|_F^2 + \mathrm{trace}(\log^2(\Lambda D^{-1}) = d^2\left((U, D), (V, \Lambda)\right). \end{aligned}$$ ∎

Proof of Theorem 3.8. We use the following lemmas.

LEMMA B.1. *For any* $D, L \in \mathrm{Diag}(p)$, $T_{bad} = \{t \in \mathbb{R} : \mathcal{J}_{D\exp(Lt)} \neq \mathcal{J}_{D,L}\}$ *has at most* $p(p-1)/2$ *elements.*

Proof. Let $D = \mathrm{diag}(d_1, \ldots, d_p)$ and $L = \mathrm{diag}(l_1, \ldots, l_p)$. For $1 \le i < j \le p$, if $i \sim_{D,L} j$, i.e., $d_i = d_j$ and $l_i = l_j$, then $i \sim_{D\exp(Lt)} j$ (or $d_i \exp(l_i t) = d_j \exp(l_j t)$) for all t and the indices i and j are in a same block of $\mathcal{J}_{D,L}$ and $\mathcal{J}_{D\exp(Lt)}$ for all t. If

$i \not\sim_{D,L} j$, then $i \sim_{D\exp(Lt)} j$ for at most one t. The result follows from the fact that there are only $p(p-1)/2$ pairs of i and j. ∎

LEMMA B.2. *Let $I \subset \mathbb{R}$ be a positive-length interval containing 0. Let $G \subset \mathrm{SO}(p)$ be a Lie subgroup of $\mathrm{SO}(p)$, and let $\mathfrak{g} \subset \mathfrak{so}(p)$ denote the Lie algebra of G. Then for any $A, B \in \mathfrak{so}(p)$,*

$$B - A \in \mathfrak{g},\ (ad_B)^j(A) \in \mathfrak{g} \ \text{ for all } \ j \geq 1 \tag{B.2}$$

if and only if there exists a C^∞ map $g : I \to G$ such that

$$\exp(tB) = \exp(tA)g(t) \ \text{ for all } \ t \in I. \tag{B.3}$$

Proof. Throughout the proof, the "prime" symbol denotes derivative, not transpose. Suppose first that (B.3) holds. For $t \in I$, define $X(t) = g(t)^{-1}g'(t)$ and $Y(t) = g(t)^{-1}Ag(t)$. Since $g'(t) \in T_{g(t)}G$, we have $X(t) = g(t)^{-1}g'(t) \in T_IG = \mathfrak{g}$. Thus, X is a C^∞ map $I \to \mathfrak{g}$, and Y is a C^∞ map $I \to \mathfrak{so}(p)$. Differentiating (B.3) gives

$$\begin{aligned}\exp(tB)B &= \exp(tA)(Ag(t) + g'(t)) = \exp(tA)g(t)(Y(t) + X(t)\\ &= \exp(tB)(Y(t) + X(t).\end{aligned}$$

Therefore,

$$X(t) + Y(t) = B \ \text{ for all } \ t \in I. \tag{B.4}$$

A simple computation leads to

$$Y'(t) = [Y(t), X(t)] = -[X(t), Y(t)]. \tag{B.5}$$

From (B.4), we get

$$X'(t) = -Y'(t) = [X(t), Y(t)] = [X(t), B - X(t)] = [X(t), B], \tag{B.6}$$

and consequently

$$X'(t) = -\mathrm{ad}_B(X(t)). \tag{B.7}$$

Equation (B.7) is a constant-coefficient linear differential equation for a function $X : I \to \mathfrak{so}(p)$. The general solution is therefore

$$X(t) = \exp(-t\mathrm{ad}_B)X(0). \tag{B.8}$$

From general Lie group theory, we have, for $W \in \mathfrak{g}$, $\exp(\mathrm{ad}_W) = \mathrm{Ad}_{\exp(W)}$, where $\mathrm{Ad}_h : \mathfrak{g} \to \mathfrak{g}$ is conjugation by h. Thus $X(t)$ can also be written as

$$X(t) = \exp(-tB)X(0)\exp(tB). \tag{B.9}$$

It is easy to check that (B.9) solves the ordinary differential equation (B.7) (or, equivalently, (B.6)). Now, since $g(0) = I$, we have $Y(0) = A$. Thus (B.4) implies that $X(0) = B - A$. Since $X(t)$ lies in the vector space $\mathfrak{g}$, so do the derivatives $X^{(j)}(0)$ of all orders $j \geq 0$. So $B - A = X(0) \in \mathfrak{g}$, $X^{(j)}(0) \in \mathfrak{g}$ for all $j \geq 1$. Hence

$$\begin{aligned}B - A &= X(0) \in \mathfrak{g},\\ (\mathrm{ad}_B)^j(A) &= -(\mathrm{ad}_B)^j(B - A) = \pm X^{(j)}(0) \in \mathfrak{g} \ \text{ for all } \ j \geq 1,\end{aligned}$$

leading to the conditions (B.2).

Next, suppose (B.2) holds. Define $C = B - A$. Then $(\mathrm{ad}_B)^j(C) = -(\mathrm{ad}_B)^j(A) \in \mathfrak{g}$ for all $j \geq 1$. Define $X(t) = \exp(-t\mathrm{ad}_B)C$; then, for all t,

$$X(t) = C + \sum_{j=1}^{\infty} \frac{1}{j!}(-t\mathrm{ad}_B)^j(C) \in \mathfrak{g}.$$

Note that X is the unique solution of the initial-value problem $X'(t) = -[B, X(t)]$, $X(0) = C$. From (B.8)–(B.9), $X(t)$ can be written as

$$X(t) = \exp(-tB)C\exp(tB) = B - \exp(-tB)C\exp(-tB).$$

It is a known fact that for a compact Lie group G with Lie algebra $\mathfrak{g}$, and any smooth $X_1 : \mathbb{R} \to \mathfrak{g}$, the initial-value problem $g(t)^{-1}g'(t) = X_1(t)$, $g(0) =$ identity element of G has a unique solution, that the solution is smooth, and that the maximal time-domain of the solution is all of $\mathbb{R}$. Since $G \subset \mathrm{SO}(p)$ is compact and X is smooth, we have a smooth solution $g : \mathbb{R} \to G$ of the initial-value problem

$$g(t)^{-1}g'(t) = X(t), \quad g(0) = I.$$

Define $Y_1(t) = g(t)^{-1}Ag(t)$. Then, as computed in (B.5), $Y_1'(t) = [Y_1(t), X(t)]$. Define $Y_2(t) = \exp(-tB)A\exp(tB) = B - X(t)$. Then $Y_2'(t) = -X'(t) = [B, X(t)] = [X(t) + Y_2(t), X(t)] = [Y_2(t), X(t)]$. Note also that $Y_1(0) = A = Y_2(0)$. Thus Y_1, Y_2 satisfy the same linear initial-value problem for a function $Y : \mathbb{R} \to \mathfrak{so}(p)$, and hence are identically equal. Therefore, $X(t) + Y_1(t) = X(t) + Y_2(t) = B$ for all t.

Define $h(t) = \exp(tA)g(t)\exp(-tB)$. Then

$$\begin{aligned} h'(t) &= \exp(tA)(Ag(t) + g'(t) - g(t)B)\exp(-tB) \\ &= \exp(tA)g(t)(Y_1(t) + X(t) - B)\exp(-tB) \equiv 0, \end{aligned}$$

so $h(t) = h(0) = I$ for all $t \in \mathbb{R}$. Therefore the function g satisfies (B.3). ∎

We now provide a proof of Theorem 3.8. Let $(V, \Lambda) \in (\mathrm{SO} \times \mathrm{Diag}^+)(p)$, $B \in \mathfrak{so}(p)$, $N \in \mathrm{Diag}(p)$, and $\gamma = \gamma(V, \Lambda, B, N) : I \to (\mathrm{SO} \times \mathrm{Diag}^+)(p)$.

Suppose first that $\chi(t) = c(\gamma(t))$ for all $t \in I$. Theorem 3.3 indicates that for all $t \in I$, there exist $R(t) \in G_{D\exp(Lt)}$ and $\pi_t \in S_p$ such that

$$\exp(tB)V = \exp(tA)UR(t)P'_{\pi_t}, \tag{B.10}$$

$$\exp(tN)\Lambda = \exp(tL_{\pi_t})D_{\pi_t}. \tag{B.11}$$

Setting $t = 0$, we have

$$V = UR(0)P'_{\pi_0}, \quad \Lambda = D_{\pi_0}. \tag{B.12}$$

From (B.11) and (B.12), we get $tN + \log D_{\pi_0} = tL_{\pi_t} + \log D_{\pi_t}$. Thus, for $i = 1, \ldots, p$ and all $t \in I\backslash\{0\}$,

$$n_i = l_{t,i} + \frac{c_{t,i}}{t}, \tag{B.13}$$

where $n_i, l_{t,i}$, and $c_{t,i}$ are the ith diagonal entries of N, L_{π_t}, and $\log(D_{\pi_t}D_{\pi_0}^{-1})$, respectively. As t, π_t, and i range over their possible values, there are only finitely many

values of $l_{t,i}$ and $c_{t,i}$. Therefore, unless $c_{t,i} = 0$ for all i and $t \neq 0$, the right-hand side of (B.13) is unbounded as $t \to 0$, contradicting the constancy of n_i. Thus,

$$D_{\pi_t} = D_{\pi_0},\ L_{\pi_t} = N = L_{\pi_0} \text{ for all } t \in I. \tag{B.14}$$

Next, let $T_{bad} = \{t \in I : \mathcal{J}_{D\exp(Lt)} \neq \mathcal{J}_{D,L}\}$, a finite set (cf. Lemma B.1). On $I\backslash T_{bad}$, the partition $\mathcal{K}$ determined by $\exp(tN)\Lambda = \exp(tL_{\pi_0})D_{\pi_0}$ is constant, as is the partition $\mathcal{J}_{D,L}$ determined by $D\exp(Lt)$. Thus we can replace π_t, for all $t \in I\backslash T_{bad}$, by a constant permutation π_∞ that carries $\mathcal{J}_{D,L}$ to $\mathcal{K}$, without affecting the truth of (B.10)–(B.11). With this replacement applied to (B.10), we have

$$\exp(tB)V = \exp(tB)UR(0)P'_{\pi_0} = \exp(tA)UR(t)P'_{\pi_\infty} \tag{B.15}$$

for $t \in I\backslash T_{bad}$. Letting $\tilde{A} = U'AU$, $\tilde{B} = U'AU$, and $P = P'_{\pi_\infty}P_{\pi_0}$, we rewrite (B.15) as

$$\exp(t\tilde{B}) = \exp(t\tilde{A})g(t) \tag{B.16}$$

for all $t \in I\backslash T_{bad}$, where $g(t) = R(t)PR(0)'$. Thus

$$R(t) = \exp(-t\tilde{A})\exp(t\tilde{B})R(0)P',\ t \in I\backslash T_{bad}. \tag{B.17}$$

The right-hand side of (B.17) is continuous on I. Hence the left-hand side of (B.17) continuously extends to each $t_{bad} \in T_{bad}$, replacing $R(t_{bad})$ with $\lim_{t\to t_{bad}} R(t)$. With this replacement for $R(t)$, (B.16) is true for all $t \in I$, and the new functions $t \mapsto R(t)$ and $t \mapsto g(t)$ are continuous on all of I. Evaluating (B.16) at $t = 0$, we get $I = g(0) = R(0)PR(0)'$, implying that $P = I$.

Note that $R(t) \in G_{\mathcal{J}_{D,L}}$ for all $t \in I\backslash T_{bad}$. Since $G_{\mathcal{J}_{D,L}}$ is a closed subset of $\mathrm{SO}(p)$, continuity implies that $R(t) \in G_{\mathcal{J}_{D,L}}$ for all $t \in T_{bad}$ as well. From (B.17), it then follows that $t \mapsto R(t)$ is a C^∞ function $I \to G_{\mathcal{J}_{D,L}}$. Therefore there exists a C^∞ map $g : I \to G_{\mathcal{J}_{D,L}}$ satisfying (B.16) for all $t \in I$. Then Lemma B.2 shows that the asserted conditions for B are necessary.

For the sufficiency, Lemma B.2 shows that there exists a C^∞ function $g : I \to G_{\mathcal{J}_{D,L}}$ satisfying (B.16). From this it is easy to check that, for R and π as in the assumption, the eigen-composition of $\gamma(URP'_\pi, D_\pi, B, L_\pi)$ is χ. □

Proof of Corollary 3.9. Let $t_0 \in I$ be such that eigenvalues of $\chi(t_0)$ are all distinct. Then $\mathcal{J}_{D,L} = \cap_{t\in I}\mathcal{J}_{D\exp(Lt)} \subset \mathcal{J}_{D\exp(Lt_0)}$. Since $\mathcal{J}_{D\exp(Lt_0)}$ has only singleton blocks, so does $\mathcal{J}_{D,L}$. Then $\mathfrak{g}_{D,L}$ consists only of $\mathbf{0}$, and conditions (i)–(ii) in Theorem 3.8 are simplified to $B = A$. Finally, Theorem 3.3 shows that any $R \in G_{D,L}$ is an even sign-change matrix. □

Theorem 3.11 is easily obtained, and we omit the proof.

Proof of Theorem 3.12. By Theorem 3.11, it suffices to show the triangle inequality. Fix a version of Z, say (U, D). As all eigenvalues of D are distinct, there exists a unique minimizer $(V_X, \Lambda_X) \in \mathcal{E}_X$ such that $d((V_X, \Lambda_X), (U, D)) \leq d((V, \Lambda), (U, D))$ for all $(V, \Lambda) \in \mathcal{E}_X$. Likewise, denote $(V_Y, \Lambda_Y) \in \mathcal{E}_Y$ the unique minimizer with respect to (U, D). Then $d_{\mathcal{SR}}(X, Z)+d_{\mathcal{SR}}(Z, Y) = d((V_X, \Lambda_X), (U, D))+d((V_Y, \Lambda_Y), (U, D))$. As d is a metric, by the triangle inequality for $((\mathrm{SO}\times\mathrm{Diag}^+)(p), d)$, $d((V_X, \Lambda_X), (U, D))+ d((V_Y, \Lambda_Y), (U, D)) \geq d((V_X, \Lambda_X), (V_Y, \Lambda_Y))$. The proof is concluded by noting that $d_{\mathcal{SR}}(X, Y) \leq d((V_X, \Lambda_X), (V_Y, \Lambda_Y))$. □

Proof of Theorem 3.15. The following lemma is used in the proof.

LEMMA B.3. *If $((U, D), (V, \Lambda))$ is minimal for $\mathcal{E}_X$ and $\mathcal{E}_Y$, then for any $R \in G_{D,\Lambda}$ and $\pi \in S_p$, the shortest-length geodesics connecting*

$$(URP'_\pi, D_\pi) \quad \textit{and} \quad (VRP'_\pi, \Lambda_\pi) \tag{B.18}$$

are also minimal. (These minimal pairs are said to be equivalent *to each other.)*

Proof of Lemma B.3. The result is obtained by two facts. (URP'_π, D_π) is a version of $X = UDU'$, and (VRP'_π, Λ_π) is a version of $Y = V\Lambda V'$; by the invariance of d (Proposition 3.7), we have $d((URP'_\pi, D_\pi), (VRP'_\pi, \Lambda_\pi)) = d((U, D), (V, \Lambda))$. □

To prove (i) of Theorem 3.15, suppose $((U_1, D_1), (V_1, \Lambda_1))$ is another minimal pair. Then there exist $R \in G_D$, $\pi \in S_p$ such that $U_1 = URP'_\pi$, $D_1 = D_\pi$. By Proposition 3.7,

$$\begin{aligned} d((U_1, D_1), (V_1, \Lambda_1)) &= d((URP'_\pi, D_\pi), (V_1, \Lambda_1)) \\ &= d((UR, D), (V_1 P'_{\pi^{-1}}, \Lambda_{1,\pi^{-1}})), \quad \text{where } \Lambda_{1,\pi^{-1}} = (\Lambda_1)_{\pi^{-1}}, \\ &= d((U, D), (V_1 P'_{\pi^{-1}} R, \Lambda_{1,\pi^{-1}})). \end{aligned}$$

The conditions on D and Λ lead to $G_D \subset G_{\Lambda_\pi}$ for all $\pi \in S_p$, which in turn leads to $R \in G_{\Lambda_{1,\pi^{-1}}}$. Since $(V_1 P'_{\pi^{-1}} R, \Lambda_{1,\pi^{-1}}) \in \mathcal{E}_Y$, the uniqueness assumption gives $V_1 = VRP_\pi$ and $\Lambda_1 = \Lambda_\pi$. Thus by Lemma B.3, all minimal pairs are equivalent to each other. Moreover, the scaling-rotation curve corresponding to the shortest-length geodesic between the minimal pair (U_1, D_1) and (V_1, Λ_1) is the same as $\chi_o(t)$ (Theorem 3.8). Thus χ_o is unique.

For (ii), let $\gamma_0(t) = \gamma(t; U, D, A, L)$ and $\gamma_1(t) = \gamma(t; U, D, B, L_1)$ be the two minimal geodesics, where $A = \log(VU')$, $L = \log(D^{-1}\Lambda)$, $B = \log(V_1 U')$, and $L_1 = \log(D^{-1}\Lambda_1)$. The length-minimizing property of minimal geodesics implies that both $\frac{d}{dt}\gamma_0(t)|_{t=0}$ and $\frac{d}{dt}\gamma_1(t)|_{t=0}$ are perpendicular to $T_{(U,D)}\mathcal{E}_X$. Note that

$$T_{(U,D)}\mathcal{E}_X = \{(UC, 0) \in \mathrm{Sym}(p) \times \mathrm{Diag}(p) : C \in \mathfrak{g}_D\}.$$

Since left and right translations by (U, D) and (R, I) are isometry of $(\mathrm{SO} \times \mathrm{Diag}^+)(p)$, we have $(T_{(U,D)}\mathcal{E}_X)^\perp = U(\mathfrak{g}_D)^\perp \oplus \mathrm{Diag}(p)$, where $U(\mathfrak{g}_D)^\perp = \{UW : W \in (\mathfrak{g}_D)^\perp \subset T_I \mathrm{SO}(p) = \mathfrak{so}(p)\}$.

Let $\tilde{A} = U'AU$, $\tilde{B} = U'BU$. Since $\frac{d}{dt}\gamma_0(t)|_{t=0} = (U\tilde{A}, DL)$ and $\frac{d}{dt}\gamma_1(t)|_{t=0} = (U\tilde{B}, DL_1)$, it follows that $\tilde{A}, \tilde{B} \in (\mathfrak{g}_D)^\perp$, and thus $\tilde{B} - \tilde{A} \in (\mathfrak{g}_D)^\perp$.

Let $\chi_0 = c \circ \gamma_0$, $\chi_1 = c \circ \gamma_1$, and assume $\chi_0 = \chi_1$. By the necessary condition (i) in Theorem 3.8, we have $\tilde{B} - \tilde{A} \in \mathfrak{g}_D$. Hence $\tilde{B} - \tilde{A} = \mathbf{0}$, implying that $B = A$ and $V_1 = V$. Then "$\chi_0 = \chi_1$" implies $L_1 = L$ as well, a contradiction. □

Proof of Theorem 4.1. (i) By (3.5), $d_{\mathcal{SR}}(X, Y) = \min_{k,j} d((U_k, D_k), (V_j, \Lambda_j))$ for $k, j = 1, 2, 3, 4$. Suppose, without loss of generality, that $(V, \Lambda) = (V_1, \Lambda_1)$. For any choice of $j = 1, 2, 3, 4$, there exist $\pi \in S_p$ and $\sigma \in \boldsymbol{\sigma}_p^+$ such that $(V_j I_\sigma P_\pi, (\Lambda_j)_\pi) = (V_1, \Lambda_1)$. Moreover, for any k, one can choose some i so that $(U_k I_\sigma P_\pi, (D_k)_\pi) = (U_i, D_i)$. Therefore, with the help of Proposition 3.7, for any k, j, there exist π, I_σ, and i satisfying

$$d((U_k, D_k), (V_j, \Lambda_j)) = d((U_k I_\sigma P_\pi, (D_k)_\pi), (V_j I_\sigma P_\pi, (\Lambda_j)_\pi)) = d((U_i, D_i), (V_1, \Lambda_1)).$$

Thus it is enough to fix a version of Y and compare the distances given by the four versions of X.

(ii) It is clear from the proof of (i) and by Proposition 3.7 that we can fix a version of Y first. Since the eigenvalues of D are identical to, say, d_1, $(U, d_1 I_2)$ is a version of X for any $U \in \mathrm{SO}(2)$. Thus choosing $U = V$ leads to the smallest distance between $U, V \in \mathrm{SO}(2)$. □

Proof of Lemma 4.2. Note that a matrix R that rotates the first two columns of

U when postmultiplied is $R = \left[\begin{smallmatrix} R_{11} & 0 \\ 0 & 1 \end{smallmatrix}\right]$, $R_{11} \in \mathrm{SO}(2)$. Using Lemma A.1(ii), we have

$$\begin{aligned}
\operatorname*{argmin}_R d((URI_\sigma P'_\pi, D_\pi), (V, \Lambda)) &= \operatorname*{argmin}_R \|\log(URI_\sigma P'_\pi V')\|_F \\
&= \operatorname*{argmin}_R \|\log(URI_\sigma P'_\pi V')\|_F = \operatorname*{argmax}_R \operatorname{trace}(URI_\sigma P'_\pi V') \\
&= \operatorname*{argmax}_R \operatorname{trace}(I_\sigma P'_\pi V' UR) = \operatorname*{argmax}_{R_{11}} \operatorname{trace}\left(\begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} R_{11} & 0 \\ 0 & 1 \end{bmatrix} \right) \\
&= \operatorname*{argmax}_{R_{11}} \operatorname{trace}(\Gamma_{11} R_{11}) + \gamma_{22}.
\end{aligned}$$

Since $R_{11} \in \mathrm{SO}(2)$, the singular values of R_{11} are unity. The result is obtained by an application of the fact from [19] that for any square matrices A and B with vectors of singular values σ_A and σ_B in non-increasing order, $|\operatorname{trace}(A'B)| \le \sigma'_A \sigma_B$. ∎

Proof of Theorem 4.3. Proofs of (i), (ii), and (iv) can be obtained by a simple extension of the proof of Theorem 4.1. For (iii), note that all versions of X and Y are $(UR_\theta I_{\sigma_k} P'_{\pi_l}, D_{\pi_l})$ and $(VR_\phi I_{\sigma_a} P'_{\pi_b}, \Lambda_{\pi_b})$. Following the lines of the proof of Theorem 4.1(i), by choosing $I_{\sigma_a} = P_{\pi_b} = I$, it is enough to compare $(UR_\theta I_{\sigma_i} P'_{\pi_j}, D_{\pi_j})$ and (VR_ϕ, Λ). Moreover, the presence of the rotation matrix R_θ allows us to restrict the choice of I_σ and π to only six pairs. For a fixed (i, j) $(i = 1, 2, 3,\ j = 1, 2)$,

$$\begin{aligned}
\min_{\theta,\phi} d((UR_\theta I_{\sigma_i} P'_{\pi_j}, D_{\pi_j}), (VR_\phi, \Lambda)) &= \min_{\theta,\phi} \left\| \log(UR_\theta I_{\sigma_i} P'_{\pi_j} R'_\phi V') \right\|_F \\
&= \max_{\theta,\phi} \operatorname{trace}(UR_\theta I_{\sigma_i} P'_{\pi_j} R'_\phi V') \qquad \text{(B.19)}
\end{aligned}$$

by Lemma A.1(ii). ∎

REFERENCES

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2009.

[2] D. C. Alexander, *Multiple-fiber reconstruction algorithms for diffusion MRI*, Ann. N. Y. Acad. Sci., 1064 (2005), pp. 113–133.

[3] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, *Geometric means in a novel vector space structure on symmetric positive-definite matrices*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 328–347.

[4] P. J. Basser, J. Mattiello, and D. LeBihan, *MR diffusion tensor spectroscopy and imaging*, Biophys. J., 66 (1994), pp. 259–267.

[5] P. G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly, *A rigorous framework for diffusion tensor calculus*, Magnet. Reson. Med., 53 (2005), pp. 221–225.

[6] S. Bonnabel and R. Sepulchre, *Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 1055–1070.

[7] O. Carmichael, J. Chen, D. Paul, and J. Peng, *Diffusion tensor smoothing through weighted Karcher means*, Electron. J. Stat., 7 (2013), pp. 1913–1956.

[8] T.-C. Chao, M.-C. Chou, P. Yang, H.-W. Chung, and M.-T. Wu, *Effects of interpolation methods in spatial normalization of diffusion tensor imaging data on group comparison of fractional anisotropy*, Magn. Reson. Imaging, 27 (2009), pp. 681–690.

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

[10] I. L. Dryden, A. Koloydenko, and D. Zhou, *Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging*, Ann. Appl. Stat., 3 (2009), pp. 1102–1123.

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

[12] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, *The generalized dynamic-factor model: Identification and estimation*, Rev. Econom. Statist., 82 (2000), pp. 540–554.

[13] J. H. Gallier, *Geometric Methods and Applications: For Computer Science and Engineering*, Texts Appl. Math. 38, Springer, New York, 2011.
[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 2012.
[15] S. Jung and X. Qiao, *A statistical approach to set classification by feature selection with applications to classification of histopathology images*, Biometrics, 70 (2014), pp. 536–545.
[16] D. Le Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat, *Diffusion tensor imaging: Concepts and applications*, J. Magn. Reson. Imaging, 13 (2001), pp. 534–546.
[17] C. Lenglet, M. Rousson, and R. Deriche, *DTI segmentation by statistical surface evolution*, IEEE Trans. Med. Imag., 25 (2006), pp. 685–700.
[18] N. Lepore, C. Brun, Y.-Y. Chou, M.-C. Chiang, R. A. Dutton, K. M. Hayashi, E. Luders, O. L. Lopez, H. J. Aizenstein, A. W. Toga, J. T. Becker, and P. M. Thompson, *Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors*, IEEE Trans. Med. Imag., 27 (2008), pp. 129–141.
[19] L. Mirsky, *A trace inequality of John von Meumann*, Monatsh. Math., 79 (1975), pp. 303–306.
[20] M. Moakher, *Means and averaging in the group of rotations*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 1–16.
[21] M. Moakher, *A differential geometric approach to the geometric mean of symmetric positive-definite matrices*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 735–747.
[22] M. Moakher and M. Zerai, *The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data*, J. Math. Imaging Vision, 40 (2011), pp. 171–187.
[23] D. Osborne, V. Patrangenaru, L. Ellingson, D. Groisser, and A. Schwartzman, *Nonparametric two-sample tests on homogeneous Riemannian manifolds, Cholesky decompositions and diffusion tensor image analysis*, J. Multivariate Anal., 119 (2013), pp. 163–175.
[24] X. Pennec, P. Fillard, and N. Ayache, *A Riemannian framework for tensor computing*, Int. J. Comput. Vis., 66 (2006), pp. 41–66.
[25] A. Schwartzman, *Random Ellipsoids and False Discovery Rates: Statistics for Diffusion Tensor Imaging Data*, Ph.D. thesis, Stanford University, Stanford, CA, 2006.
[26] A. Schwartzman, R. F. Dougherty, and J. E. Taylor, *False discovery rate analysis of brain diffusion direction maps*, Ann. Appl. Stat., 2 (2008), pp. 153–175.
[27] A. Schwartzman, R. F. Dougherty, and J. E. Taylor, *Group comparison of eigenvalues and eigenvectors of diffusion tensors*, J. Amer. Statist. Assoc., 105 (2010), pp. 588–599.
[28] A. Schwartzman, W. F. Mascarenhas, and J. E. Taylor, *Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices*, Ann. Statist., 36 (2008), pp. 2886–2919.
[29] C. G. Small, *The Statistical Theory of Shape*, Springer Ser. Statist., Springer-Verlag, New York, 1996.
[30] R. Vemulapalli and D. W. Jacobs, *Riemannian Metric Learning for Symmetric Positive Definite Matrices*, preprint, arXiv:1501.02393, 2015.
[31] Z. Wang, B. C. Vemuri, Y. Chen, and T. H. Mareci, *A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI*, IEEE Trans. Med. Imag., 23 (2004), pp. 930–939.
[32] F. Yang, Y.-M. Zhu, I. E. Magnin, J.-H. Luo, P. Croisille, and P. B. Kingsley, *Feature-based interpolation of diffusion tensor fields and application to human cardiac DT-MRI*, Med. Image Anal., 16 (2012), pp. 459–481.
[33] Y. Yuan, H. Zhu, W. Lin, and J. S. Marron, *Local polynomial regression for symmetric positive definite matrices*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 74 (2012), pp. 697–719.
[34] H. Zhu, Y. Chen, J. G. Ibrahim, Y. Li, C. Hall, and W. Lin, *Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging*, J. Amer. Statist. Assoc., 104 (2009), pp. 1203–1212.