Scaling-rotation distance and interpolation of symmetric positive-definite matrices

Sungkyu Jung † Armin Schwartzman ‡

June 16, 2014

Abstract

We introduce a new Riemannian framework for the set of symmetric positive-definite (SPD) matrices, aimed to characterize 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 transform the set of SPD matrices into a Riemannian manifold so that scaling and rotations of SPD matrices are captured by geodesics on this manifold. The problems of non-unique eigen-decompositions and eigenvalue multiplicites are addressed by finding minimal-length geodesics, which gives rise to a distance and an interpolation method for SPD matrices. Computational procedures to evaluate the minimal scaling-rotation deformations and distances are provided for the most useful cases of $2 \times 2$ and $3 \times 3$ SPD matrices. An advantage of this new geometric framework is demonstrated in its application to diffusion tensor imaging showing that the trace, determinant and fractional anisotropy of interpolated SPD matrices by minimal scaling-rotation are monotone between two extremes in many instances. This is a desirable characteristic that other Riemannian frameworks for the SPD matrices do not possess.

Keywords: symmetric positive-definite matrices, eigen decomposition, Riemannian distance, geodesics, diffusion tensors.

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
SCALING–ROTATION OF SPD MATRICES

in a biological object [4, 14, 2]. The diffusion of water is characterized by a 3D tensor, which is a $3 \times 3$ SPD matrix. The SPD matrices also appear in other contexts of tensor computing [22], tensor-based morphometry [16] and as covariance matrices in statistics. In recent years statistical analyses of SPD matrices have been received great attention [30, 24, 26, 25, 20, 29, 21].

The main challenge in the analysis of SPD matrices is that the set of $p \times p$ SPD matrices, $\text{Sym}^+(p)$, 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 [19, 22, 15, 11]. The Log-Euclidean metric, discussed in [3], is also widely used, because of its simplicity. [10] lists these popular approaches including the Cholesky decomposition-based approach of [28] and their own approach which they call the Procrustes distance. [6] proposed a different Riemannian approach for symmetric positive semidefinite matrices of fixed rank.

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. [23] proposed a scaling–rotation curve in $\text{Sym}^+(p)$, 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, called here the scaling–rotation distance, defined by the minimum amount of rotation and scaling needed to deform one SPD matrix into another.

To this aim, an alternative representation of $\text{Sym}^+(p)$, obtained by 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 non-uniqueness of the eigen-decomposition. We show that the scaling–rotation curve corresponds to geodesics in the new geometry, and 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. Precise conditions for the uniqueness of minimal curves are given.

The proposed framework not only provides a minimal deformation, but also yields a distance between SPD matrices. This distance function is a semi-metric on $\text{Sym}^+(p)$ and invariant to simultaneous rotation, scaling and inversion of SPD matrices. The invariance to 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 [19, 22, 15, 11, 3], the proposed distance is uniquely interpretable as it directly quantifies the relative scaling of eigenvalues and rotation angle between eigenvector frames of two SPD
matrices.
An additional advantage of the new geometric framework for SPD matrices is demonstrated by the interpolation obtained via the scaling–rotation curve, which exhibits regular evolution of determinant and fractional anisotropy of the interpolated SPD matrices. 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 [19, 22, 3] do not suffer from the swelling effect, which was in part the rationale to favor 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 trace 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. This is a desirable characteristic in fiber tracking of diffusion tensor imaging [5].

The proposed geometric framework for analysis of SPD matrices is viewed as an important first step to develop statistical tools for SPD matrix data that will inherit the interpretability and the advantageous regular behavior of the scaling–rotation curve. Development of tools similar to those already existing for other geometric framework, such as bi- or tri-linear interpolations [3], weighted geometric means and spatial smoothing [19, 10, 7], principal geodesic analysis [11], regression and statistical testing [30, 26, 25, 29], are much needed in the new framework. The current paper focuses to analyze minimal scaling–rotation curves and the distance defined by them.

The rest of the paper is organized as follows. The scaling–rotation curve is formally defined in Section 2. Section 3 is devoted to precisely characterize minimal scaling–rotation curves between two SPD matrices and the distance obtained accordingly. The cylindrical representation of Sym\(^+\)(p) is introduced to handle the non-uniqueness 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 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.

2 Scaling and rotation of SPD matrices
An SPD matrix \(M \in \text{Sym}^+(p)\) 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 semi-principal axes of the ellipsoid are given by eigenvector and eigenvalue pairs of \(M\). Fig. 1 illustrates some SPD matrices in 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. scaling of the eigenvalues, or stretching (shrinking) the ellipsoid along
SCALING–ROTATION OF SPD MATRICES

Denote an eigen-decomposition of $X$ by $X = U D U'$, where the columns of $U \in \text{SO}(p)$ consist of orthogonal eigenvectors of $X$, and $D \in \text{Diag}^+(p)$ is the diagonal matrix of positive eigenvalues. Here, $\text{SO}(p)$ 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 the transformation $D \exp(Lt)$ for some $L = \text{diag}(l_1, \ldots, l_p) \in \text{Diag}(p)$, $t \in \mathbb{R}$, where Diag($p$) is the set of all $p \times p$ real diagonal matrices. Each element $l_i$ of $L$ is the scaling factor for the $i$th coordinate $d_i$ of $D$. The geometric scaling is chosen to ensure $D \exp(Lt) \in \text{Diag}^+(p)$ for all $t$. As described in Appendix A, a rotation of the eigenvectors in the ambient space at a constant angular rate is described by $\exp(A t) U$, where $A \in \text{Asym}(p)$ is an antisymmetric matrix. Incorporating the scaling and rotation together results in the general scaling–rotation curve [23],

$$\chi(t) = \chi(t; U, D, A, L) = \exp(A t) U D \exp(L t) U' \exp(A' t) \in \text{Sym}^+(p), \quad t \in \mathbb{R}. \tag{1}$$

The scaling–rotation curve characterizes deformations of $X = \chi(0) \in \text{Sym}^+(p)$ so that the ellipsoid corresponding to $X$ is smoothly rotated and stretched (or shrunk) as a function of $t$. The axis and angle of rotation are characterized by the parameter $A$, while the amount of scaling is determined by the parameter $L$ (cf. Appendix A). Fig. 1 illustrates discretized trajectories of scaling–rotation curves in $\text{Sym}^+(3)$, visualized by the corresponding ellipsoids. These curves in general do not coincide with straight lines or geodesics in other geometric frameworks such as [28, 19, 22, 15, 11, 3, 10].

Given two points $X, Y \in \text{Sym}^+(p)$, 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 $Y$, $X = U D U'$, $Y = V \Lambda V'$, we could equate $\chi(1)$ and $V \Lambda V'$, and naively solve for eigenvector matrix and eigenvalue matrix separately, leading to $A = \log(V U') \in \text{Asym}(p)$, $L = \log(D^{-1} \Lambda) \in \text{Diag}(p)$. This solution is generally correct, if the eigen-decompositions of $X$ and $Y$ are chosen carefully (see Theorem 6). The difficulty is that there are many other scaling–rotation curves that also join $X$ and $Y$, due to the non-uniqueness of eigen-decomposition. Thus it is required to consider a minimal scaling–rotation curve among all such curves.
3 Minimal scaling–rotation deformations

3.1 Decomposition of SPD matrices into scaling and rotation components

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

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

We use the following notations:

**Definition 1.** For all pairs $(U, D) \in (\text{SO} \times \text{Diag}^+)(p)$ such that $X = U D U'$,

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

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

Figure 1: Scaling–rotation curves in $\text{Sym}^+(3)$: (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 understand the effect of rotation.
The many-to-one mapping $c$ from $(\text{SO} \times \text{Diag}^+(p))$ to $\text{Sym}^+(p)$ is surjective (onto). Fig. 2 illustrates the relationship between an SPD matrix and its many versions (eigen-decompositions). While $\text{Sym}^+(p)$ is an open cone, the set $(\text{SO} \times \text{Diag}^+(p))$ can be understood as the boundary of a generalized cylinder, i.e., $(\text{SO} \times \text{Diag}^+(p))$ forms a shape of cylinder whose cross-section is spherical $(\text{SO}(p))$ and the centers of the cross section are on the positive orthant of $\mathbb{R}^p$, i.e., $\text{Diag}^+(p)$. The set $(\text{SO} \times \text{Diag}^+(p))$ is a convex Riemannian manifold, as described below in Section 3.2.

Note that considering $(\text{SO} \times \text{Diag}^+(p))$ 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 not only mathematically clear but also necessary to describe the desired family of deformations. As an example, the scaling–rotation curve depicted at the middle row of Fig. 1 is made possible by allowing unordered eigenvalues.

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

**Definition 2.** For fixed $p \in \mathbb{N}$,

(i) Given a permutation $\pi$ of $p$ elements $\pi : \{1, \ldots, p\} \to \{1, \ldots, p\}$, its permutation matrix is the $p \times p$ matrix $P^+_\pi$ whose entries are all 0 except that in row $i$ the entry $\pi(i)$ equals 1. Moreover, define $P^+ = P^+_\pi$ if $\det(P^+_\pi) = 1$,

$$P^+_\pi = \begin{bmatrix} -1 & 0 \\ 0 & I_p \end{bmatrix} P^+_\pi$$

if $\det(P^+_\pi) = -1$.

(ii) A sign-change matrix is a $p \times p$ matrix $I_\sigma$ whose diagonal elements are $+1$ or $-1$ and off-diagonal elements are zero, satisfying $\det(I_\sigma) = 1$.

It is easy to check that $P^+_\pi, I_\sigma \in \text{SO}(p)$. The number of different permutation matrices $P^+_\pi$ (or sign-change matrices $I_\sigma$) is $p!$ (or $2^{p-1}$, respectively). These two types of matrices provide operations for permutation and sign-changes. In particular, for $D \in \text{Diag}{(p)}$, $D^+_\pi = P^+_\pi D P^+_\pi = \text{diag}(d^+_{\pi(1)}, \ldots, d^+_{\pi(p)}) \in \text{Diag}{(p)}$ is a diagonal matrix whose elements are permuted by $\pi$. For $U \in \text{SO}(p)$, a
column-permuted $U$, by a permutation $\pi$, is $UP'_\pi \in SO(p)$, and a sign-changed $U$, by $I_\pi$, is $U_{I_\pi} \in SO(p)$.

Introducing permutation and sign-change matrices allows us to specify all versions of an SPD matrix.

**Theorem 1.** Every version of $X = UDU'$ is of the form $(U^*, D^*) = (URP'_\pi, D_\pi)$, for $R \in SO(p)$ satisfying $RDR' = D$, permutation matrix $P'_\pi$, and $D_\pi = P_\pi DP'_\pi$. Moreover, if the eigenvalues of $X$ are all distinct, every $R \in SO(p)$ that satisfies $RDR' = D$ is a sign-change matrix.

**Remark 1.** 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 2.** 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 matrix $R$ satisfying $RDR' = D$ in Theorem 1 can be precisely characterized. For $D = \text{diag}(d_1, \ldots, d_p) \in \text{Diag}^+(p)$, let $A_D = \{(i, j) : 1 \leq i \neq j \leq p, d_i = d_j\}$ be the set of index pairs with repeated eigenvalues. Define a partition $P$ of the coordinate indices $\{1, \ldots, p\}$, where $i$ and $j$ are in a same part if $(i, j) \in A_D$, and $\{i\}$ is a singleton part if $(i, \ell) \notin A_D$ for all $1 \leq \ell \leq p$. Denote the $k$th part of $P$ by $P_k, k = 1, \ldots, \kappa$, for some $\kappa \leq p$. Let $R = (R_{ij})_{1 \leq i, j \leq p}$ be a matrix partitioned by $P$ such that for each $P_k$, the $|P_k| \times |P_k|$ submatrix $R_k = (R_{ij})_{i,j \in P_k}$ is a rotation matrix in $SO(|P_k|)$ and that $R_{ij} = 0$ if $i$ and $j$ are in different parts. Such an $R$ always satisfies the condition $RDR' = D$. Note that $R_k$ can be an orthogonal matrix in $O(|P_k|)$ as long as the resulting $R$ is in $SO(p)$. As an illustration, let $D = \text{diag}(2, 2, 1)$. Then $A_D = \{(1, 2), (2, 1)\}$ and the partition $P$ has two parts $P_1 = \{1, 2\}, P_2 = \{3\}$. The partitioned matrix $R$ is then a $3 \times 3$ block-diagonal matrix where the first $2 \times 2$ block is any $R_1 \in SO(2)$ and the last diagonal element is 1. Intuitively, $RDR'$ with this choice of $R$ behaves as if the first $2 \times 2$ block of $D, D_1$, is arbitrarily rotated. Since $D_1 = 2I_2$, rotation makes no difference.

### 3.2 A Riemannian framework for scaling and rotation of SPD matrices

The set of rotation matrices $SO(p)$ is a $p(p-1)/2$-dimensional smooth Riemannian manifold equipped with the usual Riemannian inner product for the tangent space [12, Ch. 18]. The set of positive diagonal matrices $\text{Diag}^+(p)$ is also a $p$-dimensional smooth Riemannian manifold. The set $(SO \times \text{Diag}^+)(p)$, being a direct product of two smooth and convex manifolds, is a convex Riemannian manifold [27, 1]. We state some geometric facts necessary to our discussion.

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

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


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

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

A sensible choice of Riemannian inner product at $(U, D)$ for two tangent vectors $(A_1 U, L_1 D)$ and $(A_2 U, L_2 D)$ is

$$\langle (A_1 U, L_1 D), (A_2 U, L_2 D) \rangle_{(U, D)} = \langle U'A_1 U, U'A_2 U \rangle + k\langle D^{-1} L_1 D, D^{-1} L_2 D \rangle = \frac{1}{2}\text{trace}(A_1 A_2') + k\text{trace}(L_1 L_2), \quad k > 0, \quad (2)$$

where $\langle X, Y \rangle$ for $X, Y \in \text{GL}(p)$ denotes the Frobenius inner product $\langle X, Y \rangle = \text{trace}(XY')$. We use $k = 1$ for all of our illustrations in this paper. [6] have used a structure similar to (2) for a product of a Stiefel manifold and $\text{Sym}^+(p)$. The exponential map from a tangent space $\tau(U, D)$ to $(SO \times \text{Diag}^+)(p)$ is $\text{Exp}_{(U, D)} : \tau(U, D) \to (SO \times \text{Diag}^+)(p)$,

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

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

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

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

$$\gamma(t) = \text{Exp}_{(U, D)}((AUt, LDt)). \quad (3)$$

The inner product (2) provides the geodesic distance function on $(SO \times \text{Diag}^+)(p)$. Specifically, for $(U, D)$ and $(V, \Lambda)$, the squared geodesic distance function is

$$d^2((U, D), (V, \Lambda)) = \langle (AU, LD), (AU, LD) \rangle_{(U, D)} \quad (4)$$

$$= \frac{1}{2}\|\log(VU')\|_F^2 + k\text{trace}(\log^2(\Lambda D^{-1})), \quad k > 0,$$

where $A = \log(VU')$ and $L = \log(\Lambda D^{-1})$.

The geodesic distance (4) is a proper metric, well-defined for any $(U, D)$ and $(V, \Lambda) \in (SO \times \text{Diag}^+)(p)$, 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 $(V, \Lambda)$, there are infinitely many geodesics that connect the two points, just like there are many ways of wrapping a cylinder with a string (cf. Fig. 2). There is, however, a unique minimal-length geodesic curve that connects $(U, D)$ and $(V, \Lambda)$ if $VU'$ is not an involution, i.e., if $(VU')^2 \neq I$ [18]. The rotation matrix $VU'$ is an involution if it consists of a rotation through angle $\pi$, in which case there are more than two (exactly two if $p = 2, 3$) shortest-length geodesic curves. In such
a case, \( V \) and \( U \) are said to be antipodal in \( \text{SO}(p) \), and the matrix logarithm of \( VU' \) is not unique (there is no principal logarithm), but as discussed in Appendix A \( \log(VU') \) is a solution \( A' = VU' \) whose Frobenius norm is the smallest among all such \( A \).

**Proposition 3.** The geodesic distance \((4)\) on \((\text{SO} \times \text{Diag}^+)(p)\) is invariant under simultaneous orthogonal transformation, permutation, sign changes and scaling. For any \( R_1, R_2 \in O(p) \), permutation \( \pi \) and \( S \in \text{Diag}^+(p) \), and for any \((U, D), (V, \Lambda) \in (\text{SO} \times \text{Diag}^+)(p)\), \( d((U, D), (V, \Lambda)) = d((R_1 UR_2, SD\pi), (R_1 VR_2, S\Lambda\pi))\).

### 3.3 Scaling–rotation curves as geodesics

The Riemannian manifold \((\text{SO} \times \text{Diag}^+)(p)\) equipped with the distance \((4)\) gives a precise characterization of the scaling–rotation curve. In particular, any geodesic in \((\text{SO} \times \text{Diag}^+)(p)\) corresponds to a scaling–rotation curve in \(\text{Sym}^+(p)\). The geodesic \((3)\) is identified with the scaling–rotation curve \(\chi(t) = \chi(t; U, D, A, L) \in \text{Sym}^+(p)\) \((1)\), by the eigen-composition \(\chi(t) = \chi(t)\). On the other hand, a scaling–rotation curve \(\chi(t)\) corresponds to many geodesics in \((\text{SO} \times \text{Diag}^+)(p)\).

To characterize the number of geodesics corresponding to a curve \(\chi(t)\), define

\[
A_t = \{(i, j) : 1 \leq i \neq j \leq p, d_i \exp(l_i t) = d_j \exp(l_j t), \quad (t \in \mathbb{R}),
\]

for \(D \exp(Lt) = \text{diag}(d_1 \exp(l_1 t), \ldots, d_p \exp(l_p t))\).

**Theorem 4.** Let \((U, D, A, L)\) be the parameters of a scaling–rotation curve \(\chi(t)\) in \(\text{Sym}^+(p)\).

(i) If \(\cap_{t \in \mathbb{R}} A_t = \emptyset\), then \(\chi(t)\) is identified with a family of geodesics of the form

\[
\gamma(t) = \text{Exp}(UI_{I_\sigma}D_{I_\sigma})(AU_{I_\sigma}P'_{I_\sigma}, D_{I_\sigma}L_{I_\sigma})t,
\]

for any \(I_\sigma\) and \(\pi\).

(ii) If \(\cap_{t \in \mathbb{R}} A_t \neq \emptyset\), then there exists a \(t_0\) such that \(A_{t_0} = \cap_{t \in \mathbb{R}} A_t\), and \(\chi(t)\) is identified with a family of geodesics of the form

\[
\gamma(t) = \text{Exp}(UR_{I_\sigma}D_{I_\sigma})(BU_{I_\sigma}P'_{I_\sigma}, D_{I_\sigma}L_{I_\sigma})t,
\]

for any \(I_\sigma, \pi, R \in \text{SO}(p)\) satisfying \(RD \exp(L_{t_0})R' = D \exp(L_{t_0})\), and \(B \in \text{Asym}(p)\) satisfying \((U'BU)_{ij} = (U'AU)_{ij}\) for all \((i, j) \notin A_{t_0}\).

The condition \(\cap_{t \in \mathbb{R}} A_t = \emptyset\) is satisfied if, for some \(t\), \(\chi(t)\) is an SPD matrix with distinct eigenvalues. Then \(\chi\) corresponds to only finitely many \((p!2^{p-1})\) geodesics. On the other hand, if \(\cap_{t \in \mathbb{R}} A_t\) is non-empty, then \(\chi(t)\) is invariant to particular rotations of \(U\) (given by \(UR\)) and also of infinitesimal rotation parameter \(A\). The arbitrary rotation matrix \(R\) in Theorem 4(ii) can be characterized, using the partition given by \(A_{t_0}\), as done in Remark 2. The change of coordinate \(U'AU\) expresses \(A\) in the coordinate system determined by \(U\).
3.4 Scaling–rotation distance of SPD matrices

In \((SO \times \text{Diag}^+)(p)\), consider the set of all elements whose eigen-composition is \(X\):

\[
\mathcal{O}(X) = \{(U, D) \in (SO \times \text{Diag}^+)(p) : X = UDU'\}.
\]

Since the eigen-composition is a surjective mapping, the collection of these sets \(\mathcal{O}(X)\) partitions the manifold \((SO \times \text{Diag}^+)(p)\). The use of the symbol \(\mathcal{O}\) reflects the fact that these sets are orbits, or equivalence classes. Theorem 1 above characterizes all members of \(\mathcal{O}(X)\) for any \(X\).

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

**Definition 3.** For \(X, Y \in \text{Sym}^+(p)\), the scaling–rotation distance is defined as

\[
d_{\text{SR}}(X, Y) := \inf_{(U, D) \in \mathcal{O}(X), (V, \Lambda) \in \mathcal{O}(Y)} d((U, D), (V, \Lambda)),
\]

where \(d(\cdot, \cdot)\) is the geodesic distance function (4).

The geodesic distance \(d((U, D), (V, \Lambda))\) measures the length of the shortest geodesic segment connecting \((U, D)\) and \((V, \Lambda)\). Any geodesic, mapped to \(\text{Sym}^+(p)\) by the eigen-composition, is a scaling–rotation curve connecting \(X = UDU'\) and \(Y = V\Lambda V'\). In this sense, the scaling–rotation distance \(d_{\text{SR}}\) 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 \(d_{\text{SR}}\) on \(\text{Sym}^+(p)\) is well-defined as the infimum in (7) is taken over two closed sets \(\mathcal{O}(X)\) and \(\mathcal{O}(Y)\). It has desirable invariance properties, and is a semi-metric on \(\text{Sym}^+(p)\).

**Theorem 5.** For any \(X, Y \in \text{Sym}^+(p)\), the scaling–rotation distance \(d_{\text{SR}}\) is

(i) symmetric with respect to matrix inversion, i.e., \(d_{\text{SR}}(X, Y) = d_{\text{SR}}(X^{-1}, Y^{-1})\),

(ii) invariant to simultaneous uniform scaling and rotation, i.e., \(d_{\text{SR}}(X, Y) = d_{\text{SR}}(sRXR', sRYR')\) for any \(s > 0, R \in SO(p)\),

(iii) a semi-metric on \(\text{Sym}^+(p)\). That is, \(d_{\text{SR}}(X, Y) \geq 0, d_{\text{SR}}(X, Y) = 0\) if and only if \(X = Y\), \(d_{\text{SR}}(X, Y) = d_{\text{SR}}(Y, X)\).

It is possible to find counterexamples that show the triangle inequality does not hold in general. Therefore, \(d_{\text{SR}}\) is not a proper metric on the entire set \(\text{Sym}^+(p)\).

3.5 Minimal scaling–rotation curves in \(\text{Sym}^+(p)\)

To evaluate the scaling–rotation distance (7), it is necessary to find a shortest-length geodesic in \((SO \times \text{Diag}^+)(p)\) between the orbits \(\mathcal{O}(X)\) and \(\mathcal{O}(Y)\). There are multiple geodesics connecting two orbits, because each orbit contains at least \(p!2^{n-1}\) versions (Theorem 1), as depicted in Fig. 3. We think of orbits \(\mathcal{O}(X)\)
arranged vertically in \((SO \times Diag^+)(p)\) with the mapping \(c\) (eigen-composition) as downward projections. It is clear that there exists a geodesic, say \(\gamma_h(t)\), that joins the two orbits with the minimal distance, and can be thought of as orthogonal to the orbits. As orbits were arranged vertically, this particular geodesic \(\gamma_h(t)\) is called a horizontal geodesic [1]. \((U_h, D_h)\) and \((V_h, \Lambda_h)\) are said to be horizontal if they are reached by the horizontal geodesic. The distance \(d_{\text{SR}}(X, Y)\) is the length of any horizontal geodesic segment connecting the orbits \(O(X)\) and \(O(Y)\), and is the same as \(d((U_h, D_h), (V_h, \Lambda_h))\).

**Theorem 6.** [Minimal scaling–rotation curve] Given any \(X\) and \(Y\) in \(\text{Sym}^+(p)\), a scaling–rotation curve starting at \(X\) and reaches \(Y\) at \(t = 1\),

\[
\chi_o(t; X, Y) \equiv \chi(t; U_h, D_h, A_h, L_h), \quad A_h = \log(V_h U_h^*), \quad L_h = \log(D_h^{-1} \Lambda_h),
\]

for a horizontal pair \((U_h, D_h)\) and \((V_h, \Lambda_h)\), is minimal in the sense that \(d_{\text{SR}}(X, Y) = d((U_h, D_h), (V_h, \Lambda_h)) \leq d((U, D), (V, \Lambda))\), for all \((U, D) \in O(X), (V, \Lambda) \in O(Y)\).

The above theorem tells us that for any two points \(X, Y \in \text{Sym}^+(p)\), a minimal scaling–rotation curve is determined by a horizontal pair of \(O(X)\) and \(O(Y)\). A procedure to evaluate the parameters of the minimal rotation–scaling curve and to compute the scaling–rotation distance is provided for the special cases \(p = 2, 3\) in Section 4. Note that the matrix logarithm in Theorem 6 is the principal logarithm. This is because \(U_h\) and \(V_h\) in any horizontal pair \((U_h, D_h), (V_h, \Lambda_h)\) are not antipodal in \(SO(p)\) (an application of sign-change makes the distance smaller.)
SCALING–ROTATION OF SPD MATRICES

Even though the antipodal case has been ruled out, the minimal scaling–rotation curve $\chi_o(t;X,Y)$ may not be unique. The following theorem characterizes the conditions to have a unique $\chi_o(t;X,Y)$.

**Theorem 7.** Let $(U,D)$ and $(V,\Lambda)$ be a horizontal pair of $O(X)$ and $O(Y)$, and let $\chi_o(t) = \chi(t;U,D,\log(VU'),\log(D^{-1}\Lambda))$ be the corresponding minimal scaling–rotation curve.

(i) If $(V,\Lambda)$ is the unique minimizer of $d((U,D),(V_0,\Lambda_0))$ among all $(V_0,\Lambda_0) \in O(X)$, then all horizontal geodesics between $O(X)$ and $O(Y)$ are mapped by $\epsilon$ to the unique $\chi_o(t)$ in $\text{Sym}^+(p)$.

(ii) If there exists $(V_1,\Lambda_1) \in O(X)$ such that $(V_1,\Lambda_1) \neq (V,\Lambda)$ and $(U,D),(V_1,\Lambda_1)$ are also horizontal, 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 three situations regarding the uniqueness conditions of Theorem 7: a case with unique $\chi_o$, and two non-unique cases.

**Example.** Consider $X = \text{diag}(e^{\frac{\pi}{4}},e^{-\frac{\pi}{4}})$ and $Y = R_{\theta}(2X)R'_{\theta}$, where $R_{\theta}$ is the $2 \times 2$ rotation matrix by counterclockwise angle $\theta$.

(i) If $\theta = \pi/3$, then there exists a unique minimal scaling–rotation curve between $X,Y$. This ideal case is depicted in Fig. 4, where among the four scaling–rotation curves, the red curve $\chi_4$ 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 horizontal 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 $X = \text{diag}(e^{\frac{\pi}{4}},e^{-\frac{\pi}{4}})$ and $Y = R_{\theta}X R'_{\theta}$. It is easy to see that

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

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

**4 Computation of the minimal scaling–rotation curve and scaling–rotation distance.**

We provide computation procedures for the scaling–rotation distance $d_{SR}(X,Y)$ for $X,Y \in \text{Sym}^+(2)$ or $\text{Sym}^+(3)$. Theorems 8 and 10 below provide the horizontal pair(s), based on which the exact formulation of the minimal scaling–rotation curve is evaluated in Theorem 6 above.
Figure 4: Two SPD matrices $X$ (blue) and $Y$ (green) in the cone of $\text{Sym}^+(2)$ (top left), and their four versions in a flattened $(\text{SO} \times \text{Diag}^+)(2)$ (top right). The eigen-composition of each shortest geodesic connecting versions of $X$ and $Y$ is a scaling–rotation curve in $\text{Sym}^+(2)$. Different colors represent four different such curves. The red scaling–rotation curve has the shortest geodesic distance in $(\text{SO} \times \text{Diag}^+)(2)$, and thus is minimal. Its trajectory is shown as the deformation of ellipses in the bottom panel (from leftmost $X$ to rightmost $Y$).

4.1 Scaling–rotation distance for $2 \times 2$ SPD matrices

Let $(d_1, d_2)$ be the eigenvalues of $X$, $(\lambda_1, \lambda_2)$ the eigenvalues of $Y$.

**Theorem 8.** Given any $2 \times 2$ SPD matrices $X$ and $Y$, the distance (7) 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 $(U_i, D_i), i = 1, \ldots, 4$, and for any version $(V, \Lambda)$ of $Y$,

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

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

Therefore, the minimizer of $(U_o, D_o)$ of (8) and $(V, \Lambda)$ are a horizontal pair for the case (i); $(V, D), (V, \Lambda)$ are a horizontal pair for the case (ii).
4.2 Scaling-rotation distance for $3 \times 3$ SPD matrices

Let $X, Y \in \text{Sym}^+(3)$. Let $(d_1, d_2, d_3)$ be the eigenvalues of $X$, $(\lambda_1, \lambda_2, \lambda_3)$ the eigenvalues of $Y$, without any given ordering. In order to separately analyze and catalogue all cases of eigenvalue multiplicities in Theorem 10 below, we need the following details for the case where an eigenvalue of $X$ is of multiplicity 2.

For any version $(U, D)$ with $D = \text{diag}(d_1, d_1, d_2)$, $d_1 = d_2$, all other versions of $X$ are of the form $(URI_\sigma P'_\pi, D_\sigma)$ for permutation $\pi$, sign-change matrix $I_\sigma$ and rotation matrix $R$ that rotates the first two columns of $U$ (Theorem 1). For fixed $I_\sigma$ and $\pi$, one can find the minimal rotation $\hat{R}_{\sigma, \pi}$ satisfying

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

for all such $R$, as the following lemma states.

**Lemma 9.** Let $\Gamma = I_\sigma P'_\pi V'U = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \gamma_{22} \end{bmatrix}$, where $\Gamma_{11}$ is the first $2 \times 2$ block of $\Gamma$. The minimal rotation matrix $\hat{R}_{\sigma, \pi} = \hat{R}$ is given by $\hat{R} = \begin{bmatrix} E_2E'_1 & 0 \\ 0 & 1 \end{bmatrix}$, where $E_1, E_2 \in \text{SO}(2)$ are left- and right-singular vector matrices of $\Gamma_{11} = E_3\Lambda_1E'_2$. Here, the smallest singular value of $\Lambda_1$ is negative if $\det(\Gamma_{11}) < 0$.

Each choice of $I_\sigma$ and $\pi$ produces a minimally rotated version $(\hat{U}_{\sigma, \pi}, D_\sigma) = (U\hat{R}_{\sigma, \pi}I_\sigma P'_\pi, D_\sigma)$. To provide a horizontal pair as needed in Theorem 6, a combinatorial problem involving the $3!2^{3-1} = 24$ choices of $(I_\sigma, \pi)$ needs to be solved, since the version of $X$ closest to $(V, \Lambda)$ is found by comparing distances between $(\hat{U}_{\sigma, \pi}, D_\sigma)$ and $(V, \Lambda)$. Fortunately, there are only six such minimally rotated versions corresponding to six choices of $(I_\sigma, \pi)$. In particular, we need only $\pi_1 : (1, 2, 3) \rightarrow (1, 2, 3)$, $\pi_2 : (1, 2, 3) \rightarrow (3, 1, 2)$, $\pi_3 : (1, 2, 3) \rightarrow (1, 3, 2)$, and $I_{\sigma_1} = \text{diag}(1, 1, 1)$, $I_{\sigma_2} = \text{diag}(-1, 1, -1)$, and $\hat{R}_{\sigma_j, \pi_i}$ can be found for each $(I_{\sigma_j}, \pi_i)$, $i = 1, 2, 3$, $j = 1, 2$. 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 $(V, \Lambda)$.

**Theorem 10.** Given any $3 \times 3$ SPD matrices $X$ and $Y$, the distance (7) is computed as follows.

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

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

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

where $(\hat{U}_{\sigma_j, \pi_i}, D_\sigma)$, $i = 1, 2, 3$, $j = 1, 2$ are the six minimally rotated versions.
(iii) If \( d_1 = d_2 \neq d_3 \) and \( \lambda_1 = \lambda_2 \neq \lambda_3 \), choose \( D = \text{diag}(d_1, d_2, d_3) \) and \( \Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3) \). For any versions \((U, D), (V, \Lambda)\) of \(X\) and \(Y\),

\[
d_{SR}(X,Y) = \min_{i=1,2,3, j=1,2} d((UR_{\theta_{ij}}I_\sigma P'_\pi, D_{\pi j}), (VR_{\phi_{ij}}, \Lambda)),
\]

where \( R_\theta = \exp([a]_\times) \), \( a = (0,0,\theta)' \) (cf. Appendix \(A\)), and \((\theta_{ij}, \phi_{ij})\) simultaneously maximizes \( G(\theta, \phi) = \text{trace}(UR_\theta I_\sigma P'_\pi R'_\phi V) \).

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

The minimizer \((\theta_{ij}, \phi_{ij})\) of \( G(\theta, \phi) \) in Theorem 10(iii) is found by a numerical method. Specifically, for the \(m\)th iterate \( \theta^{(m)}, \phi^{(m)} \), \( \theta^{(m+1)} \) is the solution \( \theta \) in Lemma 9, treating \( V R_{\phi^{(m)}} \) as \( V \). Likewise \( \phi^{(m+1)} \) is obtained by using Lemma 9, with the role of \( U \) and \( V \) switched. In our experiments, convergence to the unique maximum is fast and reached by only a few iterations.

5 An application to diffusion tensor computing

This work provides an interpretative geometric framework in analysis of diffusion tensor magnetic resonance images \([14]\), where diffusion tensors are given by \(3 \times 3\) SPD matrices. 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 other popular geometric frameworks such as \([22, 11, 3, 10, 6]\).

A diffusion tensor has \(p(p+1)/2 = 6\) degrees of freedom and is often summarized by mean diffusivity (MD) and fractional anisotropy (FA) \([14]\). MD measures overall diffusion intensity in a tensor \(X = VAV'\) as the average of the eigenvalues \(\text{MD}(X) = \bar{\lambda} = \text{trace}(\Lambda)/3 = (\lambda_1 + \lambda_2 + \lambda_3)/3\). FA measures the degree of anisotropy of the tensor \(X\), and is defined by

\[
\text{FA}(X) = \sqrt{\frac{3}{2} \sqrt{\frac{(\lambda_1 - \lambda)^2 + (\lambda_2 - \lambda)^2 + (\lambda_3 - \lambda)^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}}
\]

FA is zero if all eigenvalues are equal, which happens for isotropic diffusion. The highest degree of anisotropy is given by \(\text{FA} = 1\), in which case only one eigenvalue is positive and the other two approach zero.

5.1 Scaling–rotation interpolation of SPD matrices

Interpolation of tensors is important for fiber tracking, registration and spatial normalization of diffusion tensor images \([5, 8]\). In this section we provide an example where an interpolation by the minimal scaling–rotation curve is more desirable than other forms of interpolation.
Given two SPD matrices $X, Y \in \text{Sym}^+(3)$, the minimal scaling–rotation curve, $\chi_o(t; X, Y), t \in [0, 1]$, provides an interpolation of the given SPD matrices. In particular, the interpolated value at $t = t_0$ is $\chi_o(t_0) \in \text{Sym}^+(3)$. As an example, consider interpolating $X = \text{diag}(15, 2, 1)$ to $Y$, whose eigenvalues are $(100, 2, 1)$ and the principal axes of $Y$ are different from those of $X$. The first row of Fig. 5 presents the interpolated SPD matrices by $\chi_o(t; X, Y)$. Here, the left-most (right-most) ellipsoid corresponds to $X = \chi_o(0)$ (or $Y = \chi_o(1)$, respectively). This scaling–rotation interpolation traces the simultaneous rotation and scaling of $X$ toward $Y$. The interpolation is consistent with human perception when deforming $X$ to $Y$.

Other modes of interpolation are shown in Fig. 5, rows 2–4. The Euclidean interpolation is defined by $f_E(t) = tX + (1-t)Y$. The Log-Euclidean and affine-invariant Riemannian interpolations are each defined by $f_L(t) = \exp(t \log(X) + (1-t) \log(Y))$ and $f_R(t) = X^{1/2} \exp(t \log(X^{-1/2} Y X^{-1/2})) X^{1/2}$, see [3].

The Euclidean interpolation is known to suffer from the swelling effect, that is, the volume of the interpolated ellipsoids $\det(f_E(t))^{1/2}$ can be larger than the two extremes at $t = 0, 1$ [3]. This is shown in the bottom left panel of Fig. 5 for the same example. The Log-Euclidean and affine-invariant Riemannian interpolations as well as the scaling–rotation interpolation do not suffer from the swelling effect. In fact, this always holds for the scaling-rotation curve.

**Proposition 11.** For any scaling–rotation curve $\chi(t)$ joining $X = \chi(0)$ and $Y = \chi(1)$, $\det(\chi(t))$ for $t \in [0, 1]$ is monotone and lies between $\det(X)$ and $\det(Y)$.

A particularly desirable property shown in the middle panel of Fig. 5 is that FA increases monotonically. In contrast, all other interpolations of the highly anisotropic $X$ and $Y$ become less anisotropic. This phenomenon may be called a fattening effect. The scaling–rotation interpolation $\chi_o(t)$ does not suffer from the fattening effect in this example nor in most typical situations, as shown in the examples in the online supplementary material.

Moreover, the Log-Euclidean and affine-invariant Riemannian interpolations can suffer from a shrinking effect. In Fig. 5, bottom right panel, the mean diffusivity of the interpolations $f_L(t)$ and $f_R(t)$ is smaller than $\min\{\text{MD}(X), \text{MD}(Y)\}$ for some $t$. This is expected, because $f_L(t)$ and $f_R(t)$ are weighted geometric means of $X$ and $Y$. The Euclidean interpolation $f_E(t)$ is a weighted arithmetic mean, and thus $\text{MD}(f_E(t))$ is linear. Since the scaling–rotation interpolation $\chi_o(t)$ uses the geometric scaling of eigenvalues (1), the growth of $\text{MD}(\chi_o(t))$ resembles that of $\text{MD}(f_L(t))$. However, $\chi_o(t)$ shows monotone evolution in both $\text{MD}(\chi_o(t))$ and $\det(\chi_o(t))$ in the example of Fig. 5.

The advantageous regular evolution is indebted to the rotational part of the interpolation. To see this, suppose a case where the interpolation by $\chi_o(t)$ only consists of rotation (a precise example is shown in the online supplementary material). The rotational interpolation preserves all three measurements, namely the determinant, FA and MD, while the other modes of interpolations exhibit irregular behaviors in some of the measurements. This particular case of pure rotations has been reported in [5]. On the other hand, when $\chi_o(t)$ is composed
Figure 5: (Top) Interpolations of two $3 \times 3$ 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 Riemmanian interpolation. The pointy shape of ellipsoids on both ends is well-preserved in the scaling–rotation interpolation. (Bottom) Evolution of determinant, fractional anisotropy (FA) and mean diffusivity (MD) for four modes of interpolations. The scaling–rotation interpolation solely provides a monotone pattern.

of pure scaling, then $\chi_o(t) = f_L(t) = f_R(t)$ for all $t$, and there is no guarantee that MD or FA grow monotonically. But in this case all three curves behave the same. The equality of three curves in this special case is a consequence of choosing the geometric scaling of eigenvalues in the scaling–rotation curve (1) and the Riemmanian inner product (2). The example in Fig. 5 shows a case with simultaneous scaling and rotation.
In summary, while the three popular methods suffer from swelling, fattening, or shrinking effects, the scaling–rotation interpolation provides good regular evolution of all three summary statistics. More examples illustrating these effects in various scenarios are given in the online supplementary material.

A Parameterization of scaling and rotation

The matrix exponential of a square matrix \(Y\) is \(\exp(Y) = \sum_{j=0}^{\infty} \frac{Y^j}{j!}\). A unique inverse of matrix exponential, if exists, is called the principal logarithm and denoted by \(\log(X)\) [9].

The exponential map from \(\text{Diag}(p)\) to \(\text{Diag}^+(p)\), defined by the matrix exponential, is bijective. Moreover, the element-wise exponential and logarithm for the diagonal elements give the matrix exponential and logarithm for \(\text{Diag}(p)\) and \(\text{Diag}^+(p)\) [12, Ch.18].

Rotation matrices can be parameterized by antisymmetric matrices since the exponential map from \(\text{Asym}(p)\) to \(\text{SO}(p)\) is onto. Contrary to the \(\text{Diag}^+(p)\) case, the matrix exponential is not one-to-one. A unique inverse of the exponential map for a subset of \(\text{SO}(p)\) is defined by taking the principal logarithm \(A = \log(R)\). For completeness, when there exists no principal logarithm of \(R\), the notation \(\log(R)\) denotes a 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 \(R = \exp(A) \in \text{SO}(3)\) can be understood as a linear operator rotating a vector in the real 3-space by an axis \(a = (a_1, a_2, a_3)'\) with rotation angle \(\theta = \|a\|_2\) (in radians), where \(A \in \text{Asym}(3)\) is the cross-product matrix of \(a\) defined by

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

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

**Lemma 12 ([18, 12]).**

(i) (Rodrigues’ formula) Any \(A \in \text{Asym}(3)\) is parameterized by an axis \(a = \theta \hat{a} \in \mathbb{R}^3, \|\hat{a}\| = 1\), and there exists an explicit formula for the matrix exponential of \(A = [a]_\times\), \(R = \exp(A) = \mathbb{I} + [\hat{a}]_\times \sin(\theta) + [\hat{a}]^2_\times (1 - \cos(\theta))\).

(ii) For any \(R \in \text{SO}(3)\), there exists a \(\theta \in [0, \pi]\) satisfying \(2 \cos(\theta) = \text{trace}(R) - 1, \|\log(R)\|_F = \sqrt{2}\theta\). If \(\theta < \pi\), then the inverse of the matrix exponential exists and is the principal logarithm \(\log(R) = \frac{\theta}{\sin(\theta)} (R - R')\) if \(\theta \in (0, \pi)\), \(0\) if \(\theta = 0\).

If there exists no principal logarithm of \(R\), i.e., if \(\theta = \pi\), then \(\log(R)\) is defined as one of the two elements \(A \in \text{Asym}(3)\) satisfying \(\exp(A) = R\) and \(\|A\|_F = \sqrt{2\pi}\).
B Additional lemmas and proofs

Lemma 13. There exists \( t_0 \in \mathbb{R} \) satisfying \( A_{t_0} = \cap_{t \in \mathbb{R}} A_t \).

Proof. Let \( A_\infty = \cap_{t \in \mathbb{R}} A_t \). Since \( A_t \supset A_\infty \) for all \( t \), it is enough to show that there exists \( t_0 \) such that \( A_{t_0} \subset A_\infty \). First note that \( (i,j) \in A_\infty \) if and only if \( d_i = d_j, l_i = l_j \).

Define a subset of \( \mathbb{R} \), \( B_{i,j} = \{ t \in \mathbb{R} : d_i \exp(l_i t) = d_j \exp(l_j t) \} \) for \( 1 \leq i \neq j \leq p \). It is checked that if \( d_i = d_j, l_i = l_j \), then \( B_{i,j} = \mathbb{R} \); if \( d_i \neq d_j, l_i = l_j \), then \( B_{i,j} = \{ (l_i - l_j)^{-1} \log(d_i/d_j) \} \). In words, if \( (i,j) \notin A_\infty \), then the \( (i,j) \)th diagonal elements of \( D \exp(Ut) \) are equivalent at most one point. Thus \( \cup_{(i,j) \notin A_\infty} B_{i,j} \subset \mathbb{R} \) also has only finitely many elements. Choose a \( t_0 \in \mathbb{R} \setminus \cup_{(i,j) \notin A_\infty} B_{i,j} \), then for any \( (i,j) \notin A_\infty \), we have \( d_i \exp(l_i t_0) \neq d_j \exp(l_j t_0) \), that is, \( (i,j) \notin A_{t_0} \). Therefore \( A_{t_0} \subset A_\infty \) as required.

Lemma 14. Suppose \( (U,V) \) and \( (V,\Lambda) \) are horizontal. Let \( A_t = \{(i,j) : 1 \leq i \neq j \leq p, d_i (\lambda_i/d_i)^t = d_j (\lambda_j/d_j)^t \} \), \( t \in \mathbb{R} \). Then for all \( (i,j) \in \cap_{t \in \mathbb{R}} A_t \), \( (\log(UV))_{ij} = 0 \).

Proof. The case \( A_\infty = \cap_{t \in \mathbb{R}} A_t = \emptyset \) is trivial. Assume \( A_\infty \neq \emptyset \). By Theorem 6, \( \chi(t) = \chi(t;U,D,\log(VU'),\log(D^{-1}\Lambda)) \) is the minimal scaling–rotation curve, which has corresponding geodesics of the form

\[
\gamma_B(t) = \text{Exp}(U,D)(BU,D\log(D^{-1}\Lambda)t),
\]

for some \( B \in \text{Asym}(p) \) satisfying \( (U'BU)_{ij} = (U'\log(VU')U)_{ij} \) for all \( (i,j) \notin A_\infty \). In particular one can choose

\[
(U'BU)_{ij} = 0, \quad \text{for } (i,j) \in A_\infty.
\]

Note that the geodesic \( \gamma(t) = \text{Exp}(U,D)((\log(VU')U,D\log(D^{-1}\Lambda))t) \), joining \( O(UDU') \) and \( O(VAV') \) is horizontal only if the length of \( \gamma \) for \( t \in (0,1) \), \( L(\gamma : [0,1]) \), is smaller than equal to \( L(\gamma_B : [0,1]) \) (9). It can be shown that

\[
L^2(\gamma : [0,1]) - L^2(\gamma_B : [0,1]) = \sum_{(i,j) \in A_\infty} [(\log(U'V))_{ij}^2 - (U'BU)_{ij}^2] \geq 0,
\]

by (10). We get \( \sum_{(i,j) \in A_\infty} (\log(U'V))_{ij}^2 = 0 \), thus \( (\log(U'V))_{ij} = 0 \) for all \( (i,j) \in A_\infty \).

Lemma 15 (von Neumann’s trace inequality [17]). For any \( m \times n \) matrices \( A \) and \( B \) with vectors of singular values \( \sigma_A \) and \( \sigma_B \) in non-increasing order, \( |\text{trace}(A'B)| \leq \sigma'_A \sigma_B \), and the equality holds when \( A \) and \( B \) are simultaneously diagonalizable.

Proof of Theorem 1. For any \((U^*,D^*) \in (\text{SO} \times \text{Diag}^+)(p)\), there exists \( V \in \text{SO}(p) \), \( E \in \text{Diag}^+(p) \) such that \( U^* = UV \) and \( D^* = DL \). Therefore, a
version of \((U, D)\) can be written as \((UV, DL)\) satisfying \(VDLV'U' = UDU'\), or equivalently,
\[
V D L V' = D.
\] (11)

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 \(\pi\), \(DL = D_\pi\). There are at most \(p!\) possible ways to achieve this.

Observe that there exists \(R \in SO(p)\) such that \(RP'_\pi = V\), for any \(V\) and \(\pi\). (11) 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 1 can be seen from noting that \(R\) and \(D\) commute, so the eigenvector matrix of \(R\) is \(I\), with eigenvalues \(\{e^{\theta_j}, e^{-\theta_j}, 1, j = 1, \ldots, \lfloor p/2 \rfloor\}\) [13, Corollary 2.5.11]. However, all possible values of \(\theta_j\) are either 0 or \(\pi\) because \(R\) must be a real matrix. Therefore \(R\) is a sign-change matrix.

\(\square\)

**Proof of Proposition 3.** Since \(S\) commutes with other diagonal matrices,
\[
d^2 ((R_1 U R_2, SD_\pi), (R_1 V R_2, SA_\pi))
\]
\[
= \frac{1}{2} \|\log(R_1 U R_2 R'_2 V'R'_1)\|^2_F + k\text{trace}(\log^2(SA_\pi D^{-1} D_\pi S^{-1}))
\]
\[
= \frac{1}{2} \|\log(UV')\|^2_F + k\text{trace}(\log^2(\Lambda D^{-1})) = d^2 ((U, D), (V, A)).
\]

The orthogonal transformation \(R_2 \in O(p)\) includes rotations in \(SO(p)\), permutations and sign-changes since \(P'_\pi I_\pi \in SO(p) \subset O(p)\).

**Proof of Theorem 4.** It is easy to check that the eigen-composition of the geodesic (5) or (6) is \(\chi(t)\). Thus we show that for any geodesic of the form \(\gamma^*(t) = \text{Exp}_{(V, \Lambda)}((BV, \Lambda N)t)\), if the eigen-composition \(\chi^*(t) = c(\gamma^*(t))\) is the same as \(\chi(t)\), then the geodesic \(\gamma^*(t)\) is of the form (5) or (6), depending on the conditions.

One can choose \(t_0\) satisfying \(A_{t_0} = \cap_{t \in \mathbb{R}} A_t\) by Lemma 13. Thus it is sufficient to examine \(\chi(t)\) at only \(t = t_0\). Write \(\chi(t) = \exp(A(t - t_0))U_0 D_0 \exp(L(t - t_0))U'_0 \exp(A'(t - t_0))\), where \(U_0 = \exp(A_{t_0})U_0\) and \(D_0 = D_0 \exp(L_{t_0})\). Similarly, \(\chi'(t) = \exp(B(t - t_0))V_0 \Lambda_0 \exp(N(t - t_0))V'_0 \exp(B'(t - t_0))\), where \(V_0 = \exp(B_{t_0})V_0\) and \(\Lambda_0 = \Lambda \exp(N_{t_0})\). We have \(\chi^*(t) = \chi(t)\) for all \(t \in \mathbb{R}\).

From \(\chi^*(t_0) = \chi(t_0)\), we get \(V_0 \Lambda_0 V_0' = U_0 D_0 U_0'\). Theorem 1 indicates that \((V_0, \Lambda_0)\) must be of the form

\[
(V_0, \Lambda_0) = (U_0 R I_\pi P'_\pi, P_\pi D_0 P'_\pi),
\] (12)

where \(R \in SO(p)\) satisfies \(RD_0 R' = D_0\). Therefore, we have
\[
\chi^*(t) = \exp(B(t - t_0))U_0 R I_\pi P'_\pi D_0 P'_\pi \exp(N(t - t_0))P_\pi I'_\pi R'U'_0 \exp(B'(t - t_0))
\]
\[
= \exp(B(t - t_0))U_0 RD_0 \exp(N_{\pi^{-1}}(t - t_0))R'U'_0 \exp(B'(t - t_0)).
\]

Taking the derivative of \(\chi(t)\) (and \(\chi^*(t)\)) at \(t_0\) and applying a change of coordi-
nates,
\[ U_0 \frac{d}{dt} \chi(t)|_{t=t_0} U_0 = U_0' A U_0 D_0 + D_0 U_0' A' U_0 + D_0 L = A_* D_0 + D_0 A_*' + D_0 L, \]
\[ U_0' \frac{d}{dt} \chi^*(t)|_{t=t_0} U_0 = B_* D_0 + D_0 B_*' + D_0 N_{\pi-1}. \]

Since \( \frac{d}{dt} \chi^*(t) = \frac{d}{dt} \chi(t) \), we have \( A_* D_0 + D_0 A_*' = B_* D_0 + D_0 B_*' \) and \( D_0 L = D_0 N_{\pi-1} \). These terms are orthogonal under the Frobenius norm [23, sect 6.2]. The latter equation gives
\[ N = L_\pi, \quad (13) \]
regardless of whether \( \cap_{i \in \mathbb{R}} A_i = \emptyset \) or not. Since both \( A_* \) and \( B_* \) are antisymmetric, the former is then \( (A_* - B_*) D_0 = D_0 (A_* - B_*) \) or
\[ (a_{ij} - b_{ij}) d_i^0 = (a_{ij} - b_{ij}) d_j^0 \quad (14) \]
for all \( i \neq j \), where \( a_{ij}', b_{ij}' \) are the \((i,j)\)th element of \( A_* \) and \( B_* \), respectively, and \( d_i^0 \) is the \( i \)th element of \( D_0 \).

(i) If \( \cap_{i \in \mathbb{R}} A_i = \emptyset \) then every diagonal elements of \( D_0 = D \exp(L_{t_0}) \) is of multiplicity 1. Therefore, we have \( A_* = B_* \), which in turn leads to \( A = B \). Moreover, the \( R \) in (12) must be a sign-change matrix (Theorem 1). Combining \((B, N) = (A, L_\pi)\) with (12) gives \( V = U L_\pi P_\pi^* \) and \( \Lambda = D_\pi \). Thus the geodesic \( \gamma^*(t) \) has the form (5).

(ii) If \( A_{t_0} = \cap_{i \in \mathbb{R}} A_i \neq \emptyset \), then \( d_i^0 = d_j^0 \) if \( (i, j) \in A_{t_0} \), and \( d_i^0 \neq d_j^0 \) if \( (i, j) \notin A_{t_0} \). (14) then gives \( a_{ij}' = b_{ij}' \) for all \( (i, j) \notin A_{t_0} \). Since any antisymmetric matrix \( A \) and \( \exp(At) \) commute, we have \( A_* = U_0' A U_0 = \exp(U' A U_{t_0}) U' A U \exp(U' A' U_{t_0}) = U' A U \) and \( B_* = U' B U \), leading to the condition for \( B \) as \((U' A U)_{ij} = (U' B U)_{ij} \) for all \( (i, j) \notin A_{t_0} \). Combining the definitions of \( U_0, D_0, V_0, \Lambda_0 \), (12) and (13), we have \( \Lambda = D_\pi \), \( V = U \exp(-U' B U_{t_0}) \exp(U' A U_{t_0}) \exp(U' A' U_{t_0}) = D_0 \).

V is further simplified to \( V = U R L_\pi P_\pi^* \) for \( R \) with \( R D_0 R^t = D_0 \). This, together with (13), shows that the geodesic \( \gamma^*(t) \) must be of the form (6). \( \square 

Proof of Theorem 7. \) The following lemma is used in the proof.

**Lemma 16.** If \((U, D) \) and \((V, \Lambda) \) are horizontal for \( O(X) \) and \( O(Y) \), then for any \( R \in SO(p) \) satisfying both \( RDR^t = D \) and \( R\Lambda R^t = \Lambda \), permutation \( \pi \), and sign-change matrix \( I_\pi \), then the shortest-length geodesics connecting
\[ (U R P_\pi^* I_\pi, D_\pi) \quad \text{and} \quad (V R P_\pi^* I_\pi, \Lambda_\pi) \quad (15) \]
are also horizontal, and these horizontal pairs are said to be equivalent to each other.

**Proof of Lemma 16.** The result is obtained by two facts. \((U R P_\pi^* I_\pi, D_\pi) \) is a version of \( X = UDU' \) and \((V R P_\pi^* I_\pi, \Lambda_\pi) \) is a version of \( Y = VAV' \). By the
The uniqueness assumption then gives \( V \). Theorem 1 provides the precise form of \( R \) and \( L \), which leads that \((U_1, D_1)\) and \((V_1, \Lambda_1)\) are of the form (15) in Lemma 16. This shows that all geodesics by any horizontal pair \((U_1, D_1)\) and \((V_1, \Lambda_1)\) are equivalent. Moreover, the scaling–rotation curve corresponding to the shortest-length geodesic between the horizontal pair \((U_1, D_1)\) and \((V_1, \Lambda_1)\) is the same as \( \chi_o(t) \) (Theorem 4). Thus \( \chi_o \) is unique.

For (ii), note that the horizontal geodesics for \{\((U, D), (V, \Lambda)\)\} and \{\((U, D), (V_1, \Lambda_1)\)\} are not equivalent, in the sense of Lemma 16. By Theorem 6,

\[
\chi_1(t) = \chi(t; U, D, \log(VU'), \log(D^{-1}\Lambda_1)),
\]

is also minimal.

We divide the assumption \((V, \Lambda) \neq (V_1, \Lambda_1)\) into two cases.

Case I:\( \Lambda_1 \neq \Lambda \). There exists a \( t_0 \) such that the eigenvalues of \( \chi_o(t) \), or \( D \exp(\log(D^{-1}\Lambda)t_0) \), are different from those of \( \chi_1(t) \), or \( D \exp(\log(D^{-1}\Lambda_1)t_0) \). Thus \( \chi_o(t_0) \neq \chi_1(t_0) \).

Case II:\( \Lambda_1 = \Lambda \) and \( V_1 \neq V \). Theorem 1 entails that \((V_1, \Lambda_1) = (VR, \Lambda)\) for some \( R \) satisfying (i) \( RAR' = \Lambda \). Moreover, by the assumption the \( R \) also satisfies (ii) \( R \neq I \) and (iii) \( \|\log(VU')\|_F = \|\log(VRU')\|_F \). For \( \chi_1(t) = \chi(t; U, D, \log(VRU'), \log(D^{-1}\Lambda)) \), we will show \( \chi_o(t) \neq \chi_1(t) \) for some \( t \) by contradiction.

Suppose \( \chi_o(t) = \chi_1(t) \) for all \( t \). In the following, we show that \( \log(VU') = \log(VRU') \), i.e., by the uniqueness of the principal logarithm, \( R = I \). This contradicts the assumption (ii) above, and thus there exists a \( t \) such that \( \chi_o(t) \neq \chi_1(t) \).

Define \( A_t = \{(i, j) : 1 \leq i \neq j \leq p, d_i(\lambda_i/d_i)^t = d_j(\lambda_j/d_j)^t \} \), \( t \in \mathbb{R} \). If \( \bigcap_{t \in \mathbb{R}} A_t = \emptyset \), i.e., the eigenvalues of \( \chi_o(t) \) are all distinct for all \( t \) except finitely many points, then by Theorem 4(i), \( \log(VU') = \log(VRU') \).

If \( A_\infty = \bigcap_{t \in \mathbb{R}} A_t \neq \emptyset \), that is, there exists \( i \neq j \) such that \( d_i = d_j \) and \( \lambda_i = \lambda_j \), then Theorem 4(ii) further entails the condition on \( R \):

\[
(U'\log(VU')U)_{ij} = (U'\log(VRU')U)_{ij}, \quad \text{for all} \quad (i, j) \notin A_\infty.
\]

By Lemma 14, we get for all \( (i, j) \in A_\infty \),

\[
(U'\log(VU')U)_{ij} = (\log(U'V))_{ij} = 0.
\]

Since \( \Lambda_1 = \Lambda \) and \( d((U, D), (V, \Lambda)) = d((U, D), (VR, \Lambda)) \), we have \( \|\log(VU')\|_F^2 = \|\log(VRU')\|_F^2 \). Since Frobenius norm is invariant to orthogonal transformations, we get

\[
\|\log(U'V)\|_F^2 = \|\log(U'VR)\|_F^2.
\]
Combining (18) with (16) and (17) gives \( \sum_{(i,j) \in A^*} (\log(U'VR))^2 = 0 \). This with (16) gives \( \log(VU') = \log(VRU') \). □

**Proof of Theorem 8.** (i) By definition (7), \( d_{SR}(X, Y) = \min_{k,j} d((U_k, D_k), (V_j, \Lambda_j)) \), for \( k, j = 1, 2, 3, 4 \). Suppose, without loss of generality, \( (V_1, \Lambda_1) \). For any choice of \( j = 1, 2, 3, 4 \), there exist a permutation \( \pi \) and sign–change matrix \( I_\pi \) such that \( (V_\pi I_\pi, \Lambda_\pi) = (V_1, \Lambda_1) \). Moreover, for any \( k \), one can choose some \( i \) so that \( (U_k I_{\sigma_k} P_{\pi_k}, (D_k)\pi) = (U_i, D_i) \). Therefore, with a help of Proposition 3, for any \( k,j \), there exist \( \pi, I_\pi \), and \( i \) satisfying

\[
d((U_k, D_k), (V_j, \Lambda_j)) = d((U_k I_{\sigma_k} P_{\pi_k}, (D_k)\pi), (V_j I_{\sigma_j} P_{\pi_j}, (\Lambda_j)\pi)) = d((U_i, D_i), (V_1, \Lambda_1))
\]

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

(ii) It is clear from the proof of (i) and by Proposition 3 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 \( U \in SO(2) \). Thus choosing \( U = V \) leads to the smallest distance between \( U, V \in SO(2) \). □

**Proof of Lemma 9.** Note that a matrix \( R \) that rotates the first two columns of \( U \) when post-multiplied is \( R = \begin{bmatrix} R_{11} & 0 \\ 0 & 1 \end{bmatrix}, R_{11} \in SO(2) \). Using Lemma 12(ii), we have

\[
\arg\min_R d((UR_1 I_{\sigma_k} P'_{\pi_k}, D_{\pi_k}), (V_1, \Lambda_1)) = \arg\min_R \|\log(URI_{\sigma_k} P'_{\pi_k} V')\|_F \\
= \arg\min_R \|\log(URI_{\sigma_k} P'_{\pi_k} V')\|_F = \max \text{trace}(URI_{\sigma_k} P'_{\pi_k} V') \\
= \max \text{trace}(I_{\pi_k} P'_{\pi_k} V' U R_{11}) = \max \text{trace}\left( \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} R_{11} & 0 \\ 0 & 1 \end{bmatrix} \right) \\
= \max \text{trace}(\Gamma_{11} R_{11}) + \gamma_{22}.
\]

Since \( R_{11} \in SO(2) \), the singular values of \( R_{11} \) are unity. The result is obtained by an application of Lemma 15. □

**Proof of Theorem 10.** A proof of (i),(ii) and (iv) can be obtained by a simple extension of the proof of Theorem 8. For (iii), note that all versions of \( X \) and \( Y \) are \( (UR_1 I_{\sigma_k} P'_{\pi_k}, D_{\pi_k}) \) and \( (VR_1 I_{\sigma_k} P'_{\pi_k}, \Lambda_{\pi_k}) \). Following the lines of the proof of Theorem 8(i), by choosing \( I_{\pi_k} = P_{\pi_k} = I \), it is enough to compare \( (UR_1 I_{\sigma_k} P'_{\pi_k}, D_{\pi_k}) \) and \( (VR_1 I_{\sigma_k} P'_{\pi_k}, \Lambda_{\pi_k}) \). Moreover, the presence of rotation matrix \( R_{\theta} \) restricts the choice of \( I_{\sigma} \) and \( \pi \) up to only six pairs. For a fixed \( (i,j), (i = 1, 2, 3, j = 1, 2) \),

\[
\min_{\theta, \phi} d((UR_1 I_{\sigma_k} P'_{\pi_k}, D_{\pi_k}), (VR_1 I_{\sigma_k} P'_{\pi_k}, \Lambda_{\pi_k})) = \min_{\theta, \phi} \|\log(UR_1 I_{\sigma_k} P'_{\pi_k} R_{\phi} V')\|_F \\
= \max_{\theta, \phi} \text{trace}(UR_1 I_{\sigma_k} P'_{\pi_k} R_{\phi} V'), \quad (19)
\]

by Lemma 12(ii). □

**Proof of Proposition 11.** Let \( d_i \) and \( \lambda_i \) (\( i = 1, \ldots, p \)) be the eigenvalues of \( X \) and \( Y \), respectively. Denote \( d = \det(X) = \prod_{i=1}^p d_i \) and \( \lambda = \det(Y) = \prod_{i=1}^p \lambda_i \).
The determinant of any \( \chi(t) \) are \( \det(\chi(t)) = d^{1-t}\lambda^t \). Without loss of generality, if \( d \leq \lambda \), we have \( d \leq d^{1-t}\lambda^t \leq \lambda \) for all \( t \in [0,1] \), and is strictly increasing in \([0,1]\).

References

[1] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. Optimization algorithms on matrix manifolds. Princeton University Press, 2009.

[2] Daniel C Alexander. Multiple-fiber reconstruction algorithms for diffusion mri. *Annals of the New York Academy of Sciences*, 1064(1):113–133, 2005.

[3] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.*, 29(1):328–347 (electronic), 2007.

[4] Peter J. Basser, James Mattiello, and Denis LeBihan. MR Diffusion Tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.

[5] P. G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly. A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine*, 53(1):221–225, 2005.

[6] Silvère Bonnabel and Rodolphe Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM J. Matrix Anal. Appl.*, 31(3):1055–1070, 2009.

[7] Owen Carmichael, Jun Chen, Debashis Paul, and Jie Peng. Diffusion tensor smoothing through weighted Karcher means. *Electron. J. Stat.*, 7:1913–1956, 2013.

[8] Tzu-Cheng Chao, Ming-Chung Chou, Pinchen Yang, Hsiao-Wen Chung, and Ming-Ting Wu. Effects of interpolation methods in spatial normalization of diffusion tensor imaging data on group comparison of fractional anisotropy. *Magnetic Resonance Imaging*, 27(5):681 – 690, 2009.

[9] Morton L Curtis. *Matrix Groups. Universitext*. Springer Verlag., 1984.

[10] I. L. Dryden, A. Koloydenko, and D. Zhou. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*, 3:1102–1123, 2009.

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

[12] Jean H Gallier. *Geometric methods and applications: for computer science and engineering*, volume 38. Springer, 2011.
[13] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[14] Denis Le Bihan, Jean-François Mangin, Cyril Poupon, Chris A. Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabriat. Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging*, 13(4):534–46, April 2001.

[15] Christophe Lenglet, Mikaël Rousson, and Rachid Deriche. Dti segmentation by statistical surface evolution. *Medical Imaging, IEEE Transactions on*, 25(6):685–700, 2006.

[16] F Lepore, Caroline Brun, Yi-Yu Chou, Ming-Chang Chiang, Rebecca A Dutton, Kiralee M Hayashi, Eileen Luders, Oscar L Lopez, Howard J Aizenstein, Arthur W Toga, et al. Generalized tensor-based morphometry of hiv/aids using multivariate statistics on deformation tensors. *Medical Imaging, IEEE Transactions on*, 27(1):129–141, 2008.

[17] L. Mirsky. A trace inequality of john von neumann. *Monatshefte fr Mathematik*, 79(4):303–306, 1975.

[18] Maher Moakher. Means and averaging in the group of rotations. *SIAM J. Matrix Anal. Appl.*, 24(1):1–16 (electronic), 2002.

[19] Maher Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.*, 26(3):735–747 (electronic), 2005.

[20] Maher Moakher and Mourad Zerai. The riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data. *Journal of Mathematical Imaging and Vision*, 40(2):171–187, 2011.

[21] Daniel Osborne, Vic Patrangenaru, Leif Ellingson, David Groisser, and Armin Schwartzman. Nonparametric two-sample tests on homogeneous Riemannian manifolds, Cholesky decompositions and diffusion tensor image analysis. *J. Multivariate Anal.*, 119:163–175, 2013.

[22] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, 66(1):41–66, January 2006.

[23] Armin Schwartzman. *Random ellipsoids and false discovery rates: statistics for diffusion tensor imaging data*. PhD thesis, Standford University, 2006.

[24] Armin Schwartzman, Robert F. Dougherty, and Jonathan E. Taylor. False discovery rate analysis of brain diffusion direction maps. *Ann. Appl. Stat.*, 2(1):153–175, 2008.
C Supplementary material

As referenced in Section 5 of the main article, we provide visual examples on the scaling–rotation interpolation $\chi_\alpha$ of diffusion tensors, compared with the Euclidean $f_E$, Log-Euclidean $f_L$ and affine- invariant Riemannian $f_R$ interpolations. The purpose of these additional examples is to illustrate the advantageous regular evolution of determinant, fractional anisotropy and trace along the interpolated diffusion tensors.

We show interpolations from $X$ to $Y$, for five different pairs.

1. Pure rotation. Fig. 6.
2. Pure scaling. Fig. 7.
3. A moderate mix of rotation and scaling. Fig. 8.
4. A moderate mix of rotation and scaling, but with trace($X$) = trace($Y$). Fig. 9.
5. Departure from the identity. Fig. 10.
In all figures, the top four rows show various interpolations of the given $3 \times 3$ SPD matrices: Row 1–$\chi_o$, row 2–$f_E$, row 3–$f_L$ and row 4–$f_R$. The ellipsoids are colored by the direction of the first principal axis (red: left–right, green: up–down, blue: in–out), which is smoothly changed over $t$ for $\chi_o$. For other interpolations in rows 2–4, the first principal axis corresponding to the largest eigenvalue, may not be smooth (cf. Fig. 7). The bottom panel shows the evolution of determinant, fractional anisotropy (FA) and mean diffusivity (MD) for four modes of interpolations.

**Case 1:** Pure rotation.

$$X = \text{diag}(15, 5, 1),$$

$$Y = R(\frac{\pi}{3}a)\text{diag}(15, 5, 1)R(\frac{\pi}{3}a)',$$

where $R(\theta a)$ is the $3 \times 3$ rotation matrix with axis $a$ and rotation angle $\theta$ (in radians); see Appendix Lemma A.1 (i) (Rodrigues’ formula). In all examples, $a = (-0.5272, -0.6871, 0.5)'$ (normal to the screen).

![Figure 6: Pure rotation in $\chi_o$. The rotation axis is normal to the screen. The scaling–rotation interpolation solely provides a constant pattern.](image-url)
Case 2: Pure scaling.

\[ X = \text{diag}(15, 5, 1), \]
\[ Y = \text{diag}(7, 12, 8). \]

Figure 7: Pure scaling in \( \chi_0 \). It is evident that \( \chi_0 \equiv f_L \equiv f_R \). The colors of \( f_L \) and \( f_R \) are sharply changed from red to green, because the change of the principal axis corresponding to the largest eigenvalue is not smooth. On the other hand, the principal axes of \( \chi_0 \) do not change (since it is of pure scaling). Since \( \chi_0 \equiv f_L \equiv f_R \), the interpolation by \( \chi_0 \) is only as good as the interpolations by \( f_L \). This is an example where all three interpolations suffer from the fattening and shrinking effects.
**Case 3:** A moderate mix of rotation and scaling.

\[
X = \text{diag}(15, 5, 1),
Y = R(\frac{\pi}{2}a)\text{diag}(9, 12, 8)R(\frac{\pi}{2}a)'.
\]

Figure 8: A moderate mix of rotation and scaling in $\chi_o$. The scaling–rotation interpolation solely provides a *monotone* pattern.
**Case 4**: A moderate mix of rotation and scaling, but with \( \text{trace}(X) = \text{trace}(Y) \).

\[
X = R\left(-\frac{\pi}{6} a\right)\text{diag}(15, 2, 1)R\left(-\frac{\pi}{6} a\right)',
\]
\[
Y = R\left(\frac{\pi}{6} a\right)\text{diag}(10, 6, 2)R\left(\frac{\pi}{6} a\right)'.
\]

Figure 9: A moderate mix of rotation and scaling. \( \chi_o \) does not exhibit monotone (or constant) evolution of MD (or trace), but the amount of shrinking effect is much less severe than both \( f_L \) and \( f_R \), i.e. \( \text{MD}(\chi_o(t)) > \text{MD}(f_L(t)) \) for all \( t \in (0, 1) \).
Case 5: Departure from the identity.

\[ X = \text{diag}(4, 4, 4), \]
\[ Y = R\left(\frac{\pi}{3}a\right)\text{diag}(11, 11, 6)R\left(\frac{\pi}{3}a\right)'. \]

Figure 10: \( X = 4I_3 \). \( \chi_o \) is always of a pure scaling and thus \( \chi_o \equiv f_L \equiv f_R \). The color-flips in \( f_E, f_L \) and \( f_R \) are due to the non-unique eigen-decomposition and arbitrary choice of the principal axes.