STATISTICAL ANALYSIS OF RANDOM SYMMETRIC POSITIVE DEFINITE MATRICES VIA EIGEN-DECOMPOSITION

by

Brian Rooks

B.A. Mathematics, UNC Chapel Hill 2011

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial

fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Brian Rooks

It was defended on

April 4, 2018

and approved by

Sungkyu Jung, Assistant Professor Statistics

Satish Iyengar, Professor Statistics

Zhao Ren, Assistant Professor Statistics

Robert Krafty, Associate Professor Biostatistics

Dissertation Director: Sungkyu Jung, Assistant Professor Statistics

STATISTICAL ANALYSIS OF RANDOM SYMMETRIC POSITIVE DEFINITE MATRICES VIA EIGEN-DECOMPOSITION

Brian Rooks, PhD

University of Pittsburgh, 2018

The work in this dissertation is motivated by applications in the analysis of imaging data, with an emphasis on diffusion tensor imaging (DTI), a modality of MRI used to non-invasively map the structure of the brain in living subjects. In the DTI model, the local movement of water molecules within a small region of the brain is summarized by a 3-by-3 symmetric positive-definite (SPD) matrix, called a diffusion tensor. Diffusion tensors can be uniquely associated with three-dimensional ellipsoids which, when plotted, provide an image of the brain. We are interested in analyzing diffusion tensor data on the eigen-decomposition space because the eigenvalues and eigenvectors of a diffusion tensor describe the shape and orientation of its corresponding ellipsoid, respectively. One of the major contributions of this dissertation is the creation of the first statistical estimation framework for SPD matrices using the eigen-decomposition-based scaling-rotation (SR) geometric framework from Jung et al (2015). In chapter 3, we define the set of sample scaling-rotation means of a sample of SPD matrices, propose a procedure for approximating the sample SR mean set, provide conditions under which this procedure will provide a unique solution, and provide conditions guaranteeing consistency and a Central Limit Theorem for the sample SR mean set. Our procedure for approximating the sample SR mean can also be extended to compute a weighted SR mean, which can be useful for smoothing DTI data or interpolation to improve image resolution. In chapter 4, we present moment-based hypothesis tests concerning the eigenvalue multiplicity pattern of the mean of a sample of diffusion tensors which can be used to classify the mean as one of four possible shapes: isotropic, prolate, oblate, or triaxial. The derivations of these

test procedures lead to the creation of consistent estimators of the eigenvalues of the mean diffusion tensor. In the final chapter, we present a mixture distribution framework which can be used to model the variability of SPD matrices on the eigen-decomposition space, and an accompanying likelihood-based estimation procedure which can be used for estimation of parameters of interest or inference via likelihood ratio tests.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

The study of random symmetric positive definite (SPD) matrices has garnered much attention in recent years. Applications include analysis of diffusion tensor imaging (DTI) data ([1], [49], [43]), longitudinal data analysis ([14]), and estimation of covariance matrices ([8], [13]). Analysis of DTI data will be of particular interest in this thesis.

Many geometric frameworks for the space of SPD matrices have been used in the analysis of SPD matrix-valued data, including a Euclidean framework, an Affine-Invariant framework ([37]), a Log-Euclidean framework ([2]), and a Procrustes Size-and-Shape framework ([15]). Recently, a new geometric framework based on eigen-decomposition was proposed in ([28]). This minimal scaling-rotation framework identifies each SPD matrix with an ellipsoid and defines the distance between two SPD matrices as the minimal amount of rotation and scaling required to transform one ellipsoid into another. Our goal with this thesis is to develop statistical methods based on eigen-decompositions that will be useful for the analysis of DTI data.

The layout of this thesis will be as follows: Chapter 2 presents key notation, computational tools, and necessary definitions and concepts from differential geometry culminating with a description of the minimal scaling-rotation framework of [28]; Chapter 3 presents a novel location estimation framework for $Sym^{+}(p)$ based on the minimal scaling-rotation framework of [28]; Chapter 4 derives new moment-based methods for classifying the shape of population eigenvalue means of diffusion tensors; and Chapter 5 introduces methods for performing parametric mean estimation and inference for SPD matrices via eigen-decomposition.

2.0 MATHEMATICAL BACKGROUND

2.1 NOTATION AND DEFINITIONS

Let $\mathcal{M}(p)$ denote the set of $p \times p$ real-valued matrices. For any $M \in \mathcal{M}(p)$, let M^T denote the transpose of M, and let $I_p \in \mathcal{M}(p)$ denote the $p \times p$ identity matrix. The following sets of matrices will be of interest:

- $GL(p) = \{A \in \mathcal{M}(p) : det(A) \neq 0\}$ ($p \times p$ invertible matrices)
- $Sym(p) = \{A \in \mathcal{M}(p) : A = A^T\}$ ($p \times p$ symmetric matrices)
- $Sym^+(p) = \{A^T \in Sym(p) : v^T A v > 0 \text{ for all } v \in \mathbb{R}^p \setminus \{\vec{0}\}\}$ ($p \times p$ SPD matrices)
- $Diag(p) = \{A \in \mathcal{M}(p) : A_{ij} = 0 \text{ for } i \neq j\}$ ($p \times p$ diagonal matrices)
- $Diag^+(p) = \{A \in Diag(p) : A_{ii} > 0 \text{ for all } i = 1, \ldots, p\}$ ($p \times p$ positive-definite diagonal matrices)
- $\mathfrak{so}(p) = \{A \in \mathcal{M}(p) : A = -A^T\}$ ($p \times p$ skew-symmetric matrices)
- $O(p) = \{A \in \mathcal{M}(p) : AA^T = I_p\}$ ($p \times p$ orthonormal matrices)
- $SO(p) = \{A \in \mathcal{M}(p) : AA^T = I_p \text{ and } det(A) = 1\}$ ($p \times p$ rotation matrices)

Definition 2.1.1. Let $A = (a_{ij}) \in \mathcal{M}(p)$. The **Frobenius norm** $\|.\|_F$ is

$$\|A\|_F = \sqrt{tr(AA^T)} = \left(\sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij}^2 \right)^{1/2}.$$

2.2 THE MATRIX EXPONENTIAL AND LOGARITHM

Definition 2.2.1. For $A \in \mathcal{M}(p)$, the **matrix exponential** of A is

$$Exp(A) = I_p + \sum_{k=1}^{\infty} \frac{A^k}{k!}.$$

Definition 2.2.2. If there exists some $A \in \mathcal{M}(p)$ such that $Exp(A) = B$, then A is called the **matrix logarithm** of B, which we will denote as $Log(B) = A$.

The matrix exponential above is well-defined and absolutely convergent for any $A \in \mathcal{M}(p)$ [20]. If a matrix $A \in GL(p)$ has no negative real-valued eigenvalues, then there exists a unique matrix logarithm of A called the **principal matrix logarithm** which has eigenvalues belonging to the strip $\{z \in \mathbb{C} : -\pi < \mathfrak{Im}(z) < \pi\}$ [25]. Since virtually all matrices in $Diag^+(p), Sym^+(p),$ and $SO(p)$ possess no eigenvalues on the negative real line, all uses of the matrix logarithm will be referring to the principal matrix logarithm unless otherwise stated.

Our primary use of the matrix exponential and logarithm functions will be characterizing matrices from spaces with special geometric constraints by matrices from linear spaces. Computing the matrix exponential of diagonal, symmetric, and skew-symmetric matrices will be of particular interest for this thesis:

Lemma 2.2.1. *The following properties hold when the exponential matrix function is applied to the spaces $Diag(p)$, $Sym(p)$, and $\mathfrak{so}(p)$:*

1. *Let $\Lambda \in Diag(p)$ with i-th diagonal entry λ_{ii}. Then $Exp(\Lambda) \in Diag^+(p)$ with i-th diagonal entry $e^{\lambda_{ii}}$. If $D \in Diag^+(p)$ with i-th diagonal entry $d_{ii} > 0$, then $Log(D) \in Diag(p)$ with i-th diagonal entry $\log(d_{ii})$. In summary, $Exp : Diag(p) \mapsto Diag^+(p)$ is a bijective map.*
2. *If $M \in Sym(p)$ with spectral decomposition $M = UDU^T$, where $U \in SO(p)$ and $D \in Diag(p)$, then $Exp(M) = UExp(D)U^T$. Given $L \in Sym^+(p)$ with spectral decomposition $W\Lambda W^T$, where $W \in SO(p)$ and $\Lambda \in Diag^+(p)$, $Log(L) = WLog(\Lambda)W^T$. The map $Exp : Sym(p) \mapsto Sym^+(p)$ is bijective.*
3. *$Exp : \mathfrak{so}(p) \mapsto SO(p)$ is a surjective map.*

For a matrix in $Diag(p)(Diag^+(p))$, calculation of the matrix exponential (logarithm) simplifies to applying the real-valued exponential (natural logarithm) function to each diagonal entry. Similarly, for a matrix in $Sym(p)(Sym^+(p))$, calculation of the matrix exponential (logarithm) amounts to taking the exponential (natural logarithm) of each eigenvalue while leaving its corresponding eigenvector unchanged. Closed form solutions for the matrix exponential (natural logarithm) of matrices in $\mathfrak{so}(p)(SO(p))$ exist for dimensions 2 and 3, and must be computed using numerical methods for dimension $p \geq 4$.

Lemma 2.2.2. *For dimensions $p = 2$ and 3, $Exp(A)$ has the following closed form solutions for $A \in \mathfrak{so}(p)$:*

1. *For p=2: (Euler's Formula) Let $A = \theta\mathbf{J} \in \mathfrak{so}(2)$, where $\theta \in \mathbb{R}$ and*

$$\mathbf{J} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then we have

$$Exp(A) = R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

2. *For p=3: (Rodrigues' Rotation Formula) Let $\vec{a} = (a_1, a_2, a_3)^T$ with $\vec{a}^T\vec{a} = 1$ and $\theta \in \mathbb{R}$. Then we have*

$$Exp(A) = R(\theta, \vec{a}) = I_3 + \sin(\theta)[\vec{a}]_\times + (1 - \cos(\theta))([\vec{a}]_\times)^2,$$

where $A = \theta[\vec{a}]_\times$ with $[\vec{a}]_\times$ denoting the cross product matrix

$$[\vec{a}]_\times = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}.$$

Remark 2.2.1. Euler's Formula provides a rotation angle parametrization of $SO(2)$ since left multiplication of $\vec{v} \in \mathbb{R}^2$ by $R(\theta)$ rotates $\vec{v}$ counterclockwise by angle θ about the origin. Rodrigues' Rotation Formula parametrizes a 3D rotation matrix in terms of an axis and an angle since left multiplication of $\vec{v} \in \mathbb{R}^3$ by $R(\theta, \vec{a})$ rotates $\vec{v}$ by angle θ about the axis $\vec{a}$.

Remark 2.2.2. It is easy to see that $Exp : \mathfrak{so}(p) \mapsto SO(p)$ is not injective for $p = 2, 3$ since $Exp(\theta_1 \mathbf{J}) = Exp(\theta_2 \mathbf{J})$ and $Exp(\theta_1 [\vec{a}]_\times) = Exp(\theta_2 [\vec{a}]_\times)$ whenever $\theta_1 \equiv \theta_2 \pmod{2\pi}$.

Remark 2.2.3. Rodrigues' Rotation Formula is often presented with input $[\vec{b}]_\times \in \mathfrak{so}(3)$ without decomposing $\vec{b}$ into its length and direction components. For this version, simply make the substitutions $\theta = \|\vec{b}\| = \sqrt{\vec{b}^T \vec{b}}$ and $\vec{a} = \frac{\vec{b}}{\|\vec{b}\|}$.

Lemma 2.2.3. *Let $R \in SO(p)$.*

1. *p=2: Assume that $R \neq -I_2$. Then*

$$Log(R) = \begin{cases} \mathbf{0} \text{ if } R = I_2 \\ sign(r_{21}) \cos^{-1}(r_{11}) \mathbf{J} \text{ if } R \neq I_2 \end{cases}.$$

2. *$p = 3$: Assume that $\theta \not\equiv \pi \pmod{2\pi}$, where θ satisfies $tr(R) = 1 + 2\cos(\theta)$. Then*

$$Log(R) = \begin{cases} \mathbf{0} \text{ if } \theta = 0 \\ \frac{\theta}{2\sin(\theta)}(R - R^T) \text{ if } \theta \neq 0 \end{cases}.$$

Remark 2.2.4. There is not a unique solution for $Log(R)$ when $R \in SO(p)$ is an involution, meaning $R^2 = I_p$. When $p = 2$, there is only one form of an involution, $R = -I_2$, which corresponds to a rotation in the plane by angle $\pm\pi$. For $p = 3$, any involution can be characterized as a 3D rotation matrix R with $tr(R) = -1$, which corresponds to a rotation by angle $\pm\pi$ about some axis.

When $p = 2$, solutions for $Log(R)$ when R is an involution will be of the form $Log(R) = \pm\pi\mathbf{J}$. One can choose the sign that conforms with their preference for restricting θ to $[-\pi, \pi)$ or $(-\pi, \pi]$.

In the $p = 3$ case, solutions for $Log(R)$ when R is an involution will be of the form $\pm\pi[\vec{v}]_\times$, where $\vec{v}$ is the axis from the "axis-angle" parametrization of R. To solve for the axis of rotation, note for a rotation matrix $R(\theta, \vec{a})$ with rotation angle θ and axis $\vec{a}$ it holds that $R(\theta, \vec{a})\vec{a} = \vec{a}$ (i.e. $\vec{a}$ is an eigenvector of $R(\theta, \vec{a})$ with eigenvalue 1).

Lemma 2.2.4. *Let $A \in \mathcal{M}(p)$ and $t \in \mathbb{R}$. The function $c(t) = Exp(tA)$ traces a smooth curve in $GL(p)$ with directional derivative*

$$c'(t) = Exp(tA)A.$$

When $t \in [0, 1]$, the curve $c(t) = Exp(tA)$ parametrizes a smooth path on $GL(p)$ that runs from $c(0) = I_p$ to $c(1) = Exp(A)$ with "initial direction" A since $c'(0) = A$. This concept of using the matrix exponential for defining smooth curves on some matrix space is foundational in establishing the intrinsic geometric structure of curved matrix spaces of interest like $Diag^+(p), SO(p),$ and $Sym^+(p)$.

2.3 CURVES ON RIEMANNIAN MANIFOLDS

Definition 2.3.1. Let M be a differentiable manifold and assume $\epsilon > 0$. A **curve** on M passing through $q \in M$ is a smooth function $c : \mathbb{R} \mapsto M$ that satisfies $c(0) = q$.

In this thesis we will be interested in curves as matrix-valued functions. Recall that a matrix valued function $F : \mathbb{R} \mapsto \mathcal{M}(p)$ is a matrix $F(t)$ of functions $F(t) = (f_{ij}(t))$. Requiring a matrix valued function $F(t)$ to lie in a matrix space for all $t \in \mathbb{R}$ with special geometric constraints such as $Diag(p)$ or $SO(p)$ is equivalent to imposing the defining geometric characteristics of the space on the coordinate functions of $F(t)$ for all t. For example, the matrix-valued functions

$$c_1(t) = \begin{pmatrix} 2^t & 0 \\ 0 & 4^t \end{pmatrix}, c_2(t) = \begin{pmatrix} 1+t^2 & 0 \\ 0 & 1+3t^4 \end{pmatrix}, \text{and } c_3(t) = \begin{pmatrix} 1+t & 0 \\ 0 & 1+3t \end{pmatrix}$$

define smooth curves on $Diag(2)$ that pass through the identity element I_2 (i.e $c_1(0) = c_2(0) = c_3(0) = I_2$). It is easy to see that the functions $c_1(t)$ and $c_2(t)$ lie on $Diag^+(2)$ for all $t \in \mathbb{R}$, while $c_3(t)$ lies in $Diag^+(2)$ only if $t \in (-1, \infty)$.

If all of the coordinate functions of a matrix valued function F are differentiable at some t, the derivative of F at t is defined via element-wise differentiation as $F'(t) = (f'_{ij}(t))$. While

the functions $c_1(t)$, $c_2(t)$, and $c_3(t)$ all pass through I_2, they all have different trajectories at the identity

$$c_1'(0) = \begin{pmatrix} \log(2) & 0 \\ 0 & \log(4) \end{pmatrix}, c_2'(0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, c_3'(0) = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

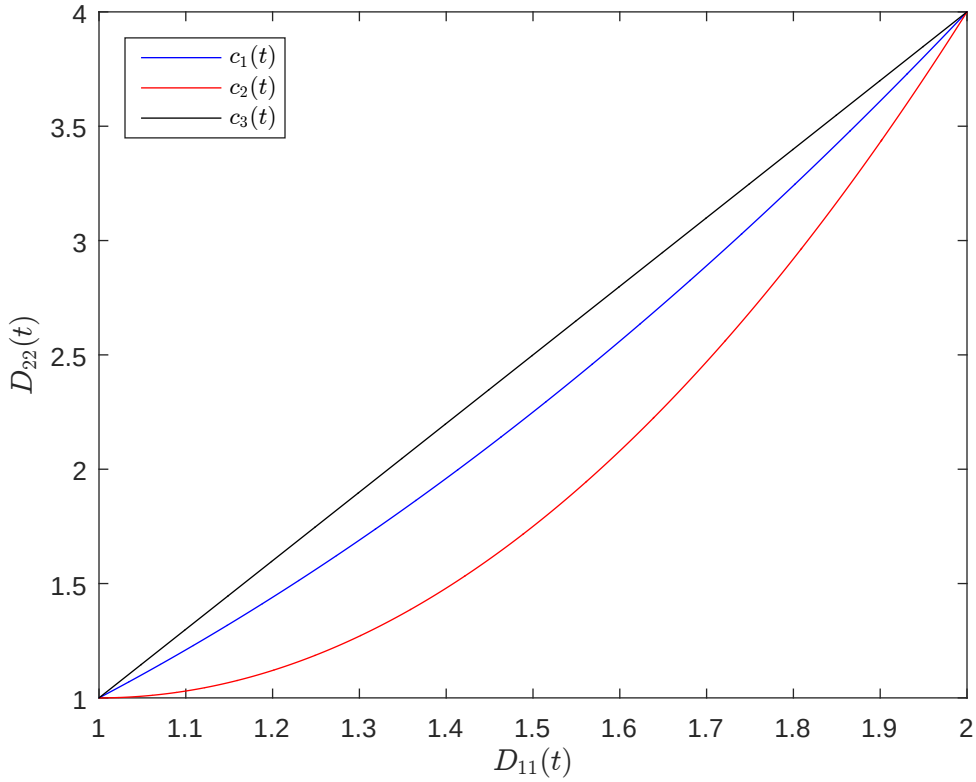

Figure 1: Plots of the diagonal functions of the curves $c_1(t), c_2(t),$ and $c_3(t)$ for $t \in [0, 1]$.

Definition 2.3.2. A **segment** connecting points $a, b \in M$ is a smooth curve on M with endpoints at a and b.

To define a measure of distance between points a and b on a smooth manifold M, one could consider the family of all segments on M connecting a and b, choose the shortest segment(s) connecting a and b (if any such minimizing curve exists), and then set the distance between points a and b equal to the length of the minimizing segment(s). To calculate the length of a segment on M, we adopt the following generalization of arc length:

Definition 2.3.3. The **arc length** of a segment $\gamma : [0,1] \mapsto M$ with endpoints $\gamma(0) = a$ and $\gamma(1) = b$ is

$$L_a^b(\gamma) = \int_0^1 \|\gamma'(t)\| dt,$$

where $\|.\|$ is some norm defined on the tangent space of M.

Remark 2.3.1. The tangent space of a manifold is defined in the next section.

Definition 2.3.4. Let $C_{a,b}$ denote the family of segments on M connecting points a and b. The **geodesic distance** between a and b, denoted as $d_G(a,b)$, equals

$$d_G(a,b) = \inf_{\gamma \in C_{a,b}} L_a^b(\gamma).$$

If such a minimizing curve exists, we will call it the **geodesic** connecting a and b.

Remark 2.3.2. A more general definition of a geodesic can be found in [16].

For $t \in [0,1]$, the curves $c_1(t), c_2(t),$ and $c_3(t)$ define segments on $Diag^+(2)$ connecting the points

$$a = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } b = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}.$$

We can metrize the tangent space of $Diag^+(2)$ with the Frobenius norm and compute the arc length of segments $c_1(t), c_2(t),$ and $c_3(t)$ as follows:

$$\begin{aligned}
L_a^b(c_1) &= \int_0^1 \|c_1'(t)\|_F dt = \int_0^1 \sqrt{(\log(2)2^t)^2 + (\log(4)4^t)^2} dt \approx 3.1628 \\
L_a^b(c_2) &= \int_0^1 \|c_2'(t)\|_F dt = \int_0^1 \sqrt{4t^2 + 144t^6} dt \approx 3.2490 \\
L_a^b(c_3) &= \int_0^1 \|c_3'(t)\|_F dt = \int_0^1 \sqrt{1^2 + 3^2} dt = \sqrt{10} \approx 3.1623.
\end{aligned}$$

From Figure 1, it is not surprising that $c_3(t)$ is the shortest of the three segments connecting a and b. In fact, $c_3(t)$ defines the geodesic connecting a and b when the Frobenius norm is used to calculate arc length. This choice of norm treats the space of all segments from a to b on $Diag^+(2)$ as flat subset of $\mathbb{R}^2$, and it is known that the shortest segment connecting two points in Euclidean space is a straight line.

The use of different norms for calculating arc lengths on manifolds will give rise to different forms of geodesics, and it can be shown that any connected Riemannian manifold

equipped with the geodesic distance function of Definition 2.3.4 is a metric space. While the existence of a minimal-length curve connecting any two points on a smooth manifold is not guaranteed in general, their existence is guaranteed for complete manifolds including $Diag^+(p)$, $SO(p)$, and $Sym^+(p)$ via the Hopf-Rinow Theorem.

2.4 THE TANGENT SPACE

Definition 2.4.1. Let M be a differentiable manifold and let C_q denote the space of all curves on M passing through $q \in M$. The **tangent space** to M at q, denoted as $T_q(M)$, can be defined as the set of all possible initial directional derivatives of curves in C_q.

Remark 2.4.1. For a more general definition of the tangent space, we refer the reader to chapter 2 of [16].

For any dimension p, it can be shown that $T_I(Diag^+(p))$ can be identified with $Diag(p)$. Any curve on $Diag^+(p)$ passing through I_p will be of the form

$$c(t) = \begin{pmatrix} f_1(t) & 0 & \cdots & 0 \\ 0 & f_2(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f_p(t) \end{pmatrix},$$

where the functions $f_1(t), \cdots, f_p(t)$ are differentiable, positive for all $t \in \mathbb{R}$, and satisfy the initial conditions $f_1(0) = \cdots = f_p(0) = 1$. Since there are no constraints on the initial derivatives $f_1'(0), \ldots, f_p'(0)$, it follows that $c'(0) \in Diag(p)$. To see why $Diag(p) \subset T_I(Diag^+(p))$, let $L \in Diag(p)$ and note that the function $\gamma(t) = Exp(Lt)$ traces a smooth curve on $Diag^+(p)$ that passes through I_p with initial direction $\gamma'(0) = L$, implying that $L \in T_I(Diag^+(p))$.

Since $Diag^+(p)$ is closed under matrix multiplication, one can construct a curve emanating from $D \in Diag^+(p)$ as $c(t)D$, where $c(t)$ is a curve on $Diag^+(p)$ that satisfies $c(0) = I_p$. Following the same steps in the previous paragraph, one will fnd that the tangent space to $Diag^+(p)$ at D is the space $T_D(Diag^+(p)) = \{LD : L \in Diag(p)\}$.

To characterize the tangent space to $SO(p)$ at the identity I_p, note that any curve $c(t)$ on $SO(p)$ passing through I_p must satisfy $c(0) = I_p$ and

$$c(t)c(t)^T = I_p. \tag{2.1}$$

Differentiating Equation 2.1 with respect to t and evaluating the derivative at 0 yields

$$c'(0)c(0)^T + c(0)c'(0)^T = c'(0) + c'(0)^T = \mathbf{0}. \tag{2.2}$$

Thus, we have that any element of $T_I(SO(p))$ must be in $\mathfrak{so}(p)$. We also have that $\mathfrak{so}(p) \subset T_I(SO(p))$ since the function $\gamma(t) = Exp(At)$, where $A \in \mathfrak{so}(p)$, defines a smooth curve on $SO(p)$ passing through I_p with initial direction $\gamma'(0) = A$.

To derive the tangent space at any $U \in SO(p)$, one can construct a curve on $SO(p)$ originating from U as $c(t)U$, where $c(t)$ is a curve on $SO(p)$ with initial condition $c(0) = I_p$, since $SO(p)$ is closed under matrix multiplication. By following the steps in the previous paragraph for characterizing $T_I(SO(p))$, one will find that $T_U(SO(p)) = \{AU : A \in \mathfrak{so}(p)\}$.

Since $Diag(p)$ and $\mathfrak{so}(p)$ are vector spaces, it is easy to see that the tangent spaces $T_D(Diag^+(p))$ and $T_U(SO(p))$ are also vector spaces. In fact, for any differentiable manifold M, it holds that $T_q(M)$ for $q \in M$ is a vector space [16]. With the tangent spaces of $Diag^+(p)$ and $SO(p)$ now well-defined, we can characterize $Diag^+(p)$ and $SO(p)$ as Riemannian manifolds:

Definition 2.4.2. A **Riemannian manifold** M of dimension p is a p-dimensional smooth manifold which is equipped with a Riemannian metric, which is a correspondence that smoothly associates to each $q \in M$ an inner product (symmetric, positive-definite, bilinear form) $\langle ., . \rangle_q$ on the tangent space $T_q(M)$.

For a Riemannian manifold M, the arc length of a curve connecting $a, b \in M$ with $\gamma(0) = a$ and $\gamma(1) = b$ can be calculated as

$$L_a^b(\gamma) = \int_0^1 \left(\langle \gamma'(t), \gamma'(t) \rangle_{\gamma(t)} \right)^{1/2} dt,$$

where $\langle ., . \rangle_{\gamma(t)}$ denotes the Riemannian metric defined on the tangent space $T_{\gamma(t)}(M)$. In the following sections we present closed form expressions for the arc lengths of geodesics defined on $Diag^+(p)$ and $SO(p)$ under special Riemannian geometric frameworks.

2.5 A RIEMANNIAN GEOMETRIC FRAMEWORK FOR DIAG-PLUS(P)

2.5.1 The Geometry of Diag-Plus(p)

The diagonal entries of any matrix in $Diag^+(p)$ lie on the positive real line, so it is easy to see that the geometric structure of $Diag^+(p)$ mirrors that of $\mathbb{R}^{p+}$, the positive orthant of $\mathbb{R}^p$. $Diag^+(p)$ is not a vector space; in fact, it is a cone since $t_1D_1 + t_2D_2 \in Diag^+(p)$ for any $D_1, D_2 \in Diag^+(p)$ and $t_1, t_2 > 0$.

$Diag^+(p)$ is a group under matrix multiplication, and is classified as a Lie group since it is a differentiable manifold whose group operation and group inverse operation are smooth. In the previous section we saw that $Diag(p)$ is the tangent space to $Diag^+(p)$ at the identity element I_p. More formally, $Diag(p)$ and $Diag^+(p)$ are linked as a Lie algebra/Lie group pair.

2.5.2 A Canonical Riemannian Framework for Diag-Plus(p)

Theorem 2.5.1. *The spaces $Diag(p)$ and $Diag^+(p)$ form a Lie Algebra/Lie Group pair. Their relationship can be characterized by the following properties:*

- *$Diag(p)$ and $Diag^+(p)$ are differentiable manifolds of dimension p.*
- *$Diag(p) = T_I(Diag^+(p))$, the tangent space of $Diag^+(p)$ at the identity element I_p.*
- *The tangent space to $Diag^+(p)$ at an arbitrary element $\Lambda \in Diag^+(p)$ is of the form $T_\Lambda(Diag^+(p)) = \{D\Lambda : D \in Diag(p)\}$.*
- *The matrix exponential with base $\Lambda \in Diag^+(p)$ from the tangent space $T_\Lambda(Diag^+(p))$ to $Diag^+(p)$ can be defined as*

$$Exp_\Lambda(L\Lambda) = Exp(L)\Lambda$$

- *The matrix logarithm with base $\Lambda \in Diag^+(p)$ from $Diag^+(p)$ to $T_\Lambda(Diag^+(p))$ can be defined as*

$$Log_\Lambda(A) = Log(A\Lambda^{-1})\Lambda$$

- *A Riemannian inner product at Λ for $L_1\Lambda, L_2\Lambda \in T_\Lambda(Diag^+(p))$ is*

$$\langle L_1\Lambda, L_2\Lambda\rangle_\Lambda := \langle \Lambda^{-1}L_1\Lambda, \Lambda^{-1}L_2\Lambda\rangle = \langle L_1, L_2\rangle = tr(L_1L_2) \tag{2.3}$$

where $\langle ., . \rangle$ denotes the Frobenius inner product $\langle X, Y\rangle = tr(XY^T)$ for $X, Y \in \mathcal{M}(p)$.

Theorem 2.5.2. *Let $\Lambda_1, \Lambda_2 \in Diag^+(p)$. The curve*

$$\gamma(t) = Exp(tLog(\Lambda_2\Lambda_1^{-1}))\Lambda_1 \tag{2.4}$$

for $t \in [0,1]$, traces the geodesic on $Diag^+(p)$ with endpoints Λ_1 and Λ_2. The length of $\gamma(.)$ using the inner product from Equation 2.3 equals the geodesic distance between Λ_1 and Λ_2 and is given by

$$\begin{aligned} d_{\mathcal{D}^+(p)}(\Lambda_1, \Lambda_2) = L_{\Lambda_1}^{\Lambda_2}(\gamma) &= \int_0^1 \left(\langle \gamma'(t), \gamma'(t)\rangle_{\gamma(t)}\right)^{1/2} dt \\ &= \left(tr(Log^2(\Lambda_2\Lambda_1^{-1}))\right)^{1/2} = \|Log(\Lambda_2\Lambda_1^{-1})\|_F. \end{aligned} \tag{2.5}$$

Remark 2.5.1. The geodesic distance function $d_{\mathcal{D}^+(p)}(.,.)$ is a proper metric on $Diag^+(p)$ and can be described as invariant under the group action since $d_{\mathcal{D}^+(p)}(X,Y)$ is not affected by simultaneous left or right multiplication for any $X, Y \in Diag^+(p)$ (i.e. $d_{\mathcal{D}^+(p)}(X,Y) = d_{\mathcal{D}^+(p)}(D_1XD_2, D_1YD_2)$ for any $D_1, D_2 \in Diag^+(p)$).

Remark 2.5.2. The distance between matrices $X, Y \in Diag^+(p)$ can be simplified into the following formula

$$\|Log(YX^{-1})\|_F = \|Log(Y) - Log(X)\|_F = \left(\sum_{i=1}^{p} (\log(y_{ii}) - \log(x_{ii}))^2\right)^{1/2}.$$

Remark 2.5.3. Since $Exp_\Lambda : T_\Lambda(Diag^+(p)) \mapsto Diag^+(p)$ is a bijective map, the tangent space to $Diag^+(p)$ provides a global linearization of the curved space $Diag^+(p)$.

2.6 A RIEMANNIAN GEOMETRIC FRAMEWORK FOR SO(P)

2.6.1 The Geometry of SO(p)

Any $R \in SO(p)$ must satisfy $R^T R = I_p$ and $det(R) = 1$. Let $R_1, \ldots, R_p$ denote the columns of a rotation matrix R. The right-most column R_p is completely determined by columns $R_1, \ldots, R_{p-1}$ since R_p must be orthogonal to the other columns and hence belongs to the one-dimensional orthogonal complement of the span of columns $R_1, \ldots, R_{p-1}$. The sign of R_p is chosen so that $det(R) = 1$. The mutual orthogonality and unit length constraints impose $\binom{p-1}{2} + p - 1 = \frac{p(p-1)}{2}$ non-linear restrictions on the remaining columns, leaving $\frac{p(p-1)}{2}$ degrees of freedom among $R_1, \ldots, R_{p-1}$. Thus, $SO(p)$ is a $\frac{p(p-1)}{2}$ dimensional non-linear space.

The identification of $R \in SO(p)$ with its first $p-1$ columns provides a characterization of $SO(p)$ as a special case of a Stiefel manifold. Recall that the Stiefel manifold $V(m,n)$ consists of all $m \times n$ matrices (with $n \leq m$) containing orthonormal columns, so $SO(p)$ can be represented as the Stiefel manifold $V(p, p-1)$. For more information on Stiefel manifolds, we refer the reader to [11].

When $p = 2$, every 2×2 rotation matrix has the representation

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix},$$

which identifies $SO(2)$ with S^1, the unit circle in $\mathbb{R}^2$. Left multiplication of $v \in \mathbb{R}^2$ by $R(\theta)$ rotates v counterclockwise about the origin by angle θ. $SO(2)$ is commutative under matrix multiplication since $R(\theta)R(\phi) = R(\theta+\phi) = R(\phi+\theta) = R(\phi)R(\theta)$, and is special since $SO(p)$ ceases to be a commutative group for $p > 2$.

For $p = 3$, any 3×3 rotation matrix has the axis-angle representation

$$R(\theta, \vec{a}) = I_3 + \sin(\theta)[\vec{a}]_\times + (1 - \cos(\theta))[\vec{a}]_\times^2$$

where $\vec{a} \in \mathbb{R}^3$, $\vec{a}^T\vec{a} = 1$, and $[\vec{a}]_\times$ denotes the cross product matrix which computes $[\vec{a}]_\times \vec{v} = \vec{a} \times \vec{v}$ for any $\vec{v} \in \mathbb{R}^3$. Left multiplication of $v \in \mathbb{R}^3$ by $R(\theta, \vec{a})$ rotates v about the axis $\vec{a}$ by angle θ. This axis-angle representation is not unique if θ is allowed to vary over all of $\mathbb{R}$ since $R(-\theta, -\vec{a}) = R(\theta, \vec{a})$ and $R(\theta_1, \vec{a}) = R(\theta_2, \vec{a})$ whenever $\theta_1 \equiv \theta_2 \pmod{2\pi}$. To obtain

a unique axis-angle parametrization for a 3D rotation matrix, it makes sense to restrict the angle of rotation $[0, \pi)$ since $R(\theta, \vec{a}) = R(2\pi - \theta, -\vec{a})$. When $\theta = \pi$, it is impossible to obtain a unique axis-angle representation since $R(\pi, \vec{a}) = R(\pi, -\vec{a})$. A rotation matrix with angle $\theta \in [0, \pi)$ and axis $\vec{a} \in S^2$ can be identified with the vector $\theta\vec{a}$, which lies in the interior of a sphere of radius π centered at the origin $(0, 0, 0)^T$, which we will denote as V_π. Rotations about the axis $\vec{a}$ by angle π can be identified with the antipodal points $\pi\vec{a}$ and $-\pi\vec{a}$ that lie on the boundary of V_π.

$SO(p)$ is a group under matrix multiplication, and is classified as a Lie group since it is a differentiable manifold whose group operation and group inverse operation are smooth. In the previous section we saw that $\mathfrak{so}(p)$ is the tangent space to $SO(p)$ at the identity element I_p. More formally, $\mathfrak{so}(p)$ and $SO(p)$ are linked as a Lie algebra/Lie group pair.

2.6.2 A Canonical Riemannian Framework for SO(p)

Theorem 2.6.1. *The spaces $\mathfrak{so}(p)$ and $SO(p)$ form a Lie Algebra/Lie Group pair. Specifically, their relationship can be characterized by the following properties:*

- *$\mathfrak{so}(p)$ and $SO(p)$ are differentiable manifolds of dimension $\frac{1}{2}p(p-1)$.*
- *$\mathfrak{so}(p) = T_I(SO(p))$, the tangent space of $SO(p)$ at the identity element I_p.*
- *The tangent space to $SO(p)$ at $U \in SO(p)$ is $T_U(SO(p)) = \{AU : A \in \mathfrak{so}(p)\}$.*
- *The matrix exponential with base $U \in SO(p)$ from the tangent space $T_U(SO(p))$ to $SO(p)$ can be defined as*

$$Exp_U(AU) = Exp(A)U$$

- *The matrix logarithm with base $U \in SO(p)$ from $SO(p)$ to $T_U(SO(p))$ can be defined as*

$$Log_U(R) = Log(RU^T)U$$

- *A Riemannian inner product at U for $A_1U, A_2U \in T_U(SO(p))$ is*

$$\langle A_1U, A_2U\rangle_U := \frac{1}{2}\langle A_1U, A_2U\rangle = \frac{1}{2}\langle A_1, A_2\rangle = \frac{1}{2}tr(A_1A_2^T) \tag{2.6}$$

where $\langle ., .\rangle$ denotes the Frobenius inner product $\langle X, Y\rangle = tr(XY^T)$ for $X, Y \in \mathcal{M}(p)$.

Theorem 2.6.2. *Let $R_1, R_2 \in SO(p)$. The curve*

$$\gamma(t) = Exp(tLog(R_2 R_1^T))R_1 \tag{2.7}$$

for $t \in [0,1]$ traces a geodesic on $SO(p)$ with endpoints R_1 and R_2. The length of $\gamma(.)$ using the inner product from Equation 2.6 equals the geodesic distance between R_1 and R_2, and is given by

$$\begin{aligned} d_{SO(p)}(R_1, R_2) = L_{R_1}^{R_2}(\gamma) &= \int_0^1 \left(\langle \gamma'(t), \gamma'(t) \rangle_{\gamma(t)}\right)^{1/2} dt \\ &= \left(\frac{1}{2} tr(Log(R_2 R_1^T) Log(R_2 R_1^T)^T)\right)^{1/2} = \frac{1}{\sqrt{2}} \|Log(R_2 R_1^T)\|_F. \end{aligned} \tag{2.8}$$

Remark 2.6.1. The geodesic distance function $d_{SO(p)}(.,.)$ is a proper metric on $SO(p)$ and is invariant under the group action since $d_{SO(p)}(X, Y)$ is invariant under simultaneous left or right multiplication for any $X, Y \in SO(p)$ (i.e. $d_{SO(p)}(X, Y) = d_{SO(p)}(U_1 X U_2, U_1 Y U_2)$ for any $U_1, U_2 \in SO(p)$).

Remark 2.6.2. For p equal to 2 or 3, $d_{SO(p)}(R_1, R_2)$ has a physical interpretation as the magnitude of a rotation angle. When identifying elements $R_1, R_2 \in SO(2)$ with points on the unit circle in $\mathbb{R}^2$, the distance $d_{SO(p)}(R_1, R_2)$ equals length of the shortest arc on the unit circle connecting the points corresponding to R_1 and R_2, which is depicted by the blue connecting arc in the left diagram of Figure 2. For 3D rotation matrices R_1 and R_2 in $SO(3)$, the distance $d_{SO(p)}(R_1, R_2)$ equals the magnitude of the rotation angle from the axis-angle representation of the rotation matrix $R_2 R_1^T$.

Remark 2.6.3. There will be multiple geodesics connecting points R_1 and R_2 whenever the rotation matrix $R_2 R_1^T$ is an involution, meaning $(R_2 R_1^T)^2 = I_p$ and $R_2 R_1^T \neq I_p$. For $p =$ 2 or 3, the matrix $R_2 R_1^T$ will be an involution if and only if $d_{SO(p)}(R_1, R_2) = \pi$. Rotation matrices in $SO(2)$ or $SO(3)$ that differ by angle π are often referred to as antipodal points. This case is illustrated for $SO(2)$ in the right diagram of Figure 2.

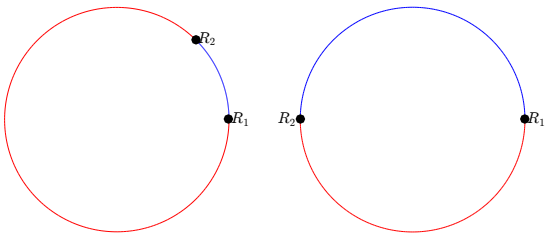

Figure 2: Visualization of rotation matrices $R_1, R_2 \in SO(2)$ as points on the unit circle.

2.7 GEOMETRIC FRAMEWORKS FOR SYM-PLUS(P)

2.7.1 The Geometry of Sym-Plus(p)

Any $M \in Sym^+(p)$ must satisfy $M = M^T$ and $x^T M x > 0$ for any $x \neq \mathbf{0}$. The symmetric condition reduces the number of degrees of freedom for any SPD matrix to $\frac{p(p+1)}{2}$. The positive definite restriction prevents $Sym^+(p)$ from being a vector space; moreover, $Sym^+(p)$ is a cone since $c_1 M_1 + c_2 M_2 \in Sym^+(p)$ for any $M_1, M_2 \in Sym^+(p)$ provided that $c_1, c_2 > 0$. As a simple illustration of $Sym^+(p)$, we consider the case when $p = 2$. Any $M \in Sym^+(2)$ must be of the form

$$M = \begin{pmatrix} a & c \\ c & b \end{pmatrix},$$

where $a > 0$, $b > 0$, and $ab - c^2 > 0$. Then $Sym^+(2)$ can be characterized as the set $\{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 > 0, x_2 > 0, x_1 x_2 - x_3^2 > 0\}$, which forms the interior of a cone in $\mathbb{R}^3$.

In addition, any $M \in Sym^+(p)$ can be identified with the ellipsoid in $\mathbb{R}^p$ whose surface coordinates $x \in \mathbb{R}^p$ satisfy $x^T M^{-1} x = 1$. In particular, the semi-principal axes of this ellipsoid and their squared lengths correspond to the eigenvectors and respective eigenvalues of M. For example, the matrix

$$X = \begin{pmatrix} 6 & 4 \\ 4 & 6 \end{pmatrix}$$

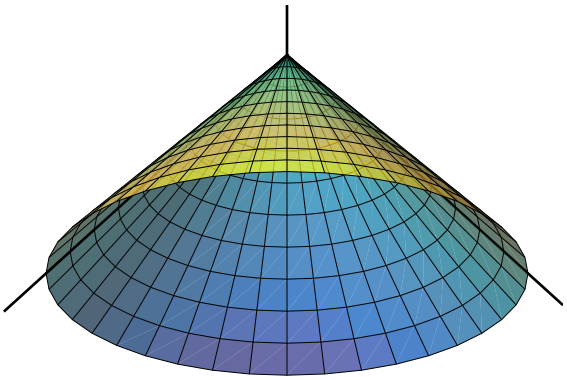

Figure 3: $Sym^+(2)$ as a cone in $\mathbb{R}^3$.

has eigenvalues 10 and 2, with respective eigenvectors $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ and $\left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$. Solutions to the equation $v^T X^{-1} v = 1$ trace an ellipse in $\mathbb{R}^2$ having principal axes of length $\sqrt{10}$ and directions $\pm\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$, and minor axes of length $\sqrt{2}$ with directions $\pm\left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$.

Unlike $Diag^+(p)$ and $SO(p)$, $Sym^+(p)$ is not closed under matrix multiplication. However, movement throughout $Sym^+(p)$ can be accomplished via the group action $\phi : GL(p) \times Sym^+(p) \mapsto Sym^+(p)$ defined as $\phi_G(M) = GMG^T \in Sym^+(p)$, where $G \in GL(p)$ and $M \in Sym^+(p)$. This conjugation action can be used to motivate an affine-invariant Riemannian geometric framework for $Sym^+(p)$ (see [18]).

2.8 A SCALING-ROTATION FRAMEWORK FOR SYM-PLUS(P)

2.8.1 Scaling-Rotation Curves

In this section we provide motivation for the scaling-rotation framework for $Sym^+(p)$ proposed in [28].

Recall that any $X \in Sym^+(p)$ has an eigen-decomposition $X = UDU^T$, where $U \in SO(p)$ and $D \in Diag^+(p)$. For any $X_1, X_2 \in Sym^+(p)$, deformation from X_1 to X_2 can be described by a combination of scaling the eigenvalues and rotating the eigenvectors of X_1. When X_1 and X_2 are viewed as ellipsoids, deformation from X_1 to X_2 occurs as a combination of rotating the axes of X_1 to align with those of X_2 and scaling the lengths of

the axes of X_1 to match the length of the axes of X_2. If X_1 and X_2 have eigen-decompositions $X_1 = U_1 D_1 U_1^T$ and $X_2 = U_2 D_2 U_2^T$ with $D_1, D_2 \in Diag^+(p)$ and $U_1, U_2 \in SO(p)$, one method for smoothly scaling the eigenvalues of X_1 to match those of X_2 and smoothly rotating the eigenvectors of X_1 to match those of X_2 is via the geodesic curves from Equation 2.4 and Equation 2.7

$$\chi_S(t) = Exp(Log(D_2 D_1^{-1})t)D_1$$

$$\chi_R(t) = Exp(Log(U_2 U_1^T)t)U_1$$

defined for $t \in [0, 1]$. The curves above can be combined via eigen-composition to create the following smooth curve in $Sym^+(p)$ running from X_1 to X_2

$$\chi_{SR}(t) = \chi_R(t)\chi_S(t)\chi_R(t)^T,$$

which we will refer to as a scaling-rotation curve.

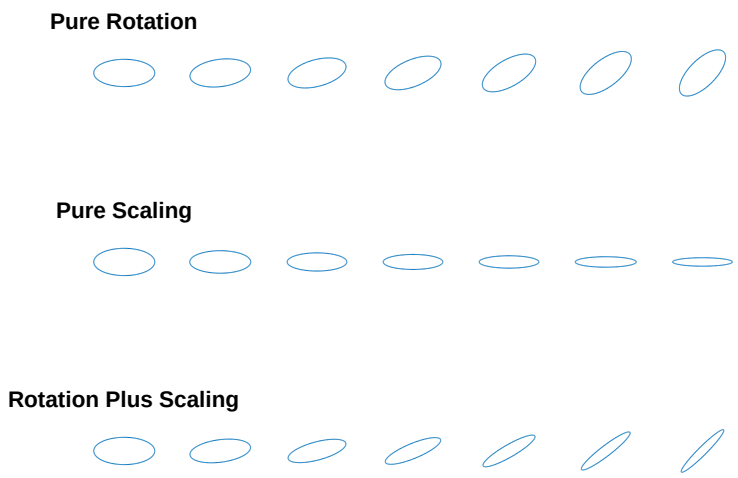

Figure 4: Interpolation between 2×2 SPD matrices along various scaling-rotation curves.

One can define the distance between X_1 and X_2 as the length of a scaling-rotation curve connecting X_1 and X_2. Developing a scaling-rotation framework for $Sym^+(p)$ will require a geometric framework for $Diag^+(p) \times SO(p)$, the cross product space which contains all possible eigen-decompositions, and a method to account for the non-uniqueness of eigen-decomposition for a given SPD matrix.

2.8.2 A Riemannian Framework for the Eigen-Decomposition Space

To define a scaling-rotation framework for $Sym^+(p)$, we first describe a Riemannian geometric framework for $Diag^+(p) \times SO(p)$ that combines the frameworks for $Diag^+(p)$ and $SO(p)$ that were described in section 2.5 and section 2.6.

Theorem 2.8.1. *The spaces $Diag(p) \times \mathfrak{so}(p)$ and $Diag^+(p) \times SO(p)$ form a Lie Algebra/Lie Group pair. Their relationship can be characterized by the following properties:*

- *$Diag(p) \times \mathfrak{so}(p)$ and $Diag^+(p) \times SO(p)$ are differentiable manifolds of dimension $p + \frac{p(p-1)}{2}$.*
- *$Diag(p) \times \mathfrak{so}(p) = T_I(Diag^+(p) \times SO(p))$, the tangent space to $Diag^+(p) \times SO(p)$ at the identity element (I_p, I_p).*
- *The tangent space to $Diag^+(p) \times SO(p)$ at $(\Lambda, U) \in Diag^+(p) \times SO(p)$ is of the form $T_{(\Lambda,U)}(Diag^+(p) \times SO(p)) = \{(L\Lambda, AU) : L \in Diag(p) \text{ and } A \in \mathfrak{so}(p)\}$.*
- *The matrix exponential with base $(\Lambda, U) \in Diag^+(p) \times SO(p)$ from the tangent space $T_{(\Lambda,U)}(Diag^+(p) \times SO(p))$ to $Diag^+(p) \times SO(p)$ can be defined as*

$$Exp_{(\Lambda,U)}((L\Lambda, AU)) = (Exp(L)\Lambda, Exp(A)U).$$

- *The matrix logarithm with base $(\Lambda, U) \in Diag^+(p) \times SO(p)$ from $Diag^+(p) \times SO(p)$ to $T_{(\Lambda,U)}(Diag^+(p) \times SO(p))$ can be defined as*

$$Log_{(\Lambda,U)}(D, R) = (Log(D\Lambda^{-1})\Lambda, Log(RU^T)U).$$

- *A Riemannian inner product at (Λ, U) for $(L_1\Lambda, A_1U), (L_2\Lambda, A_2U) \in T_{(\Lambda,U)}(Diag^+(p) \times SO(p))$ is*

$$\begin{aligned} \langle (L_1\Lambda, A_1U), (L_2\Lambda, A_2U) \rangle_{(\Lambda,U)} &= \langle \Lambda^{-1}L_1\Lambda, \Lambda^{-1}L_2\Lambda \rangle + \frac{k}{2}\langle A_1U, A_2U \rangle \\ &= tr(L_1L_2) + \frac{k}{2}tr(A_1A_2^T), k > 0, \end{aligned} \tag{2.9}$$

where $\langle ., . \rangle$ denotes the Frobenius inner product $\langle X, Y \rangle = tr(XY^T)$.

Theorem 2.8.2. *Let* $(\Lambda_1, R_1), (\Lambda_2, R_2) \in Diag^+(p) \times SO(p)$. *The curve*

$$\gamma(t) = (Exp(Log(\Lambda_2\Lambda_1^{-1})t)\Lambda_1, Exp(Log(R_2R_1^T)t)R_1) \tag{2.10}$$

for $t \in [0,1]$ *traces a geodesic on* $Diag^+(p) \times SO(p)$ *with endpoints at* (Λ_1, R_1) *and* (Λ_2, R_2). *The length of* $\gamma(.)$ *using inner product Equation 2.9 equals the geodesic distance between* (Λ_1, R_1) *and* (Λ_2, R_2), *and is given by*

$$\begin{aligned} d_{SR}((\Lambda_1, R_1), (\Lambda_2, R_2)) = L_{(\Lambda_1,R_1)}^{(\Lambda_2,R_2)}(\gamma) &= \int_0^1 (\langle \gamma'(t), \gamma'(t) \rangle_{\gamma(t)})^{1/2} dt \\ &= \left(tr(Log^2(\Lambda_2\Lambda_1^{-1})) + \frac{k}{2} tr(Log(R_2R_1^T) Log(R_2R_1^T)^T) \right)^{1/2} \\ &= \left(\|Log(\Lambda_2\Lambda_1^{-1})\|_F^2 + \frac{k}{2} \|Log(R_2R_1^T)\|_F^2 \right)^{1/2}. \end{aligned} \tag{2.11}$$

Remark 2.8.1. When $k = 1$, $d_{SR}(.,.)$ is equal to $(d^2_{\mathcal{D}^+(p)}(.,.) + d^2_{SO(p)}(.,.))^{1/2}$, where $d^+_{\mathcal{D}}(p)(.,.)$ and $d_{SO(p)}(.,.)$ are the distance functions defined in Equation 2.5 and Equation 2.8.

Remark 2.8.2. The distance function above inherits the invariance properties of Equation 2.5 and Equation 2.8.

2.8.3 Non-uniqueness of Eigen-decomposition

Given $X \in Sym^+(p)$ with eigen-decomposition $X = UDU^T$, the eigen-decomposition (D, U) is not unique since it is always possible to find a matrix $R \in SO(p)$, $R \neq I_p$, such that $R^TDR \in Diag^+(p)$ and (R^TDR, UR) is also an eigen-decomposition of X.

For example, the matrix

$$X = \begin{pmatrix} 6 & 4 \\ 4 & 6 \end{pmatrix}$$

has eigen-decompositions (D_1, U_1) and (D_2, U_2) with

$$D_1 = \begin{pmatrix} 10 & 0 \\ 0 & 2 \end{pmatrix} U_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$D_2 = \begin{pmatrix} 2 & 0 \\ 0 & 10 \end{pmatrix} U_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

since $U_1 D_1 U_1^T = X = U_2 D_2 U_2^T$.

If (D_1, U_1) and (D_2, U_2) are eigen-decompositions of some $X \in Sym^+(p)$, we will call them versions of the eigen-decomposition of X. Different versions of the eigen-decomposition of an SPD matrix differ by special orthonormal transformations, and under special circumstances, there will be finitely many versions. In order to describe the transformations that link the different versions of an eigen-decomposition, we recall some special types of matrices:

Definition 2.8.1. Let π denote a permutation of $\{1, 2, \ldots, p\}$. The **permutation matrix** associated with π, which we will denote as P_π, is a $p \times p$ matrix whose entries are all zero except for elements $(i, \pi(i))$ for $i = 1, \ldots, p$, which equal 1.

For any $A \in \mathcal{M}(p)$, left multiplication of A by P_π permutes the rows of A according to π, while right multiplication by P_π^T permutes the columns of A according to π. Any permutation matrix P_π satisfies $P_\pi P_\pi^T = I_p$ and $det(P_\pi) = \pm 1$, where the sign of the determinant equals the sign of the permutation π. To create permutation matrices with strictly positive determinants, we define the modified permutation matrix

$$P_\pi^+ = \begin{cases} P_\pi & \text{if } det(P_\pi) = 1 \\ P_\pi A_+ & \text{if } det(P_\pi) = -1 \end{cases},$$

where

$$A_+ = \begin{pmatrix} -1 & \mathbf{0} \\ \mathbf{0} & I_{p-1} \end{pmatrix}.$$

It is straightforward to show that $P_\pi^+ \in SO(p)$ for any permutation π.

Definition 2.8.2. A **sign change** matrix, which we denote as I_σ, is a $p \times p$ matrix whose diagonal elements equal 1 or -1 and off-diagonal elements equal zero, with the restriction that $det(I_\sigma) = 1$.

Let $\{i_1, \ldots, i_k\} \subset \{1, \ldots, p\}$ and suppose I_σ is a sign-change matrix with -1 on the diagonal of columns $i_1, \ldots, i_k$. For any $A \in \mathcal{M}(p)$, left multiplication of A by I_σ changes the signs of rows $i_1, \ldots, i_k$ of A, while right multiplication of A by I_σ changes the signs of columns $i_1, \ldots, i_k$ of A. For any sign change matrix I_σ, it holds that $I_\sigma I_\sigma^T = I_p$ and $det(I_\sigma) = 1$, implying that $I_\sigma \in SO(p)$.

Lemma 2.8.1. *The following results hold for modified permutation matrices and sign change matrices:*

1. *There are $p!$ distinct modified permutation matrices of size $p \times p$.*
2. *There are 2^{p-1} distinct sign change matrices of size $p \times p$.*
3. *The spaces of modified permutation matrices and sign change matrices are subsets of $SO(p)$ and are closed under matrix multiplication.*

Theorem 2.8.3. *Let $X \in Sym^+(p)$. Every eigen-decomposition $X = UDU^T$, where $U \in SO(p)$ and $D \in Diag^+(p)$, is of the form $(D^*, U^*) = (D_\pi, UR(P_\pi^+)^T)$, where $R \in SO(p)$ satisfies $RDR^T = D$ and $D_\pi = P_\pi^+ D (P_\pi^+)^T$. If the eigenvalues of X are distinct, then every $R \in SO(p)$ that satisfies $RDR^T = D$ is a sign change matrix.*

Corollary 2.8.1. *Let $X \in Sym^+(p)$. If the eigenvalues of X are distinct, then there are $2^{p-1}p!$ different version of the eigen-decomposition of X.*

Returning to our example

$$X = \begin{pmatrix} 6 & 4 \\ 4 & 6 \end{pmatrix},$$

we now know that X also has eigen-decompositions (D_3, U_3) and (D_4, U_4) with

$$D_3 = \begin{pmatrix} 10 & 0 \\ 0 & 2 \end{pmatrix} U_3 = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

$$D_4 = \begin{pmatrix} 2 & 0 \\ 0 & 10 \end{pmatrix} U_4 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

in addition to its previously mentioned decompositions (D_1, U_1) and (D_2, U_2). All four eigen-decompositions of X are represented in Figure 5. For a given decomposition, red (black) arrows follow the direction of the first (second) column of the eigenvector matrix and have length equal to the square root of the first (second) eigenvalue.

Note that when an SPD matrix has some eigenvalues that are equal, there will be infinitely many eigen-decompositions. For example, the 2×2 SPD matrix $X = 4I_2$ can be decomposed as $X = U(4I_2)U^T$, where U is any member of $SO(2)$.

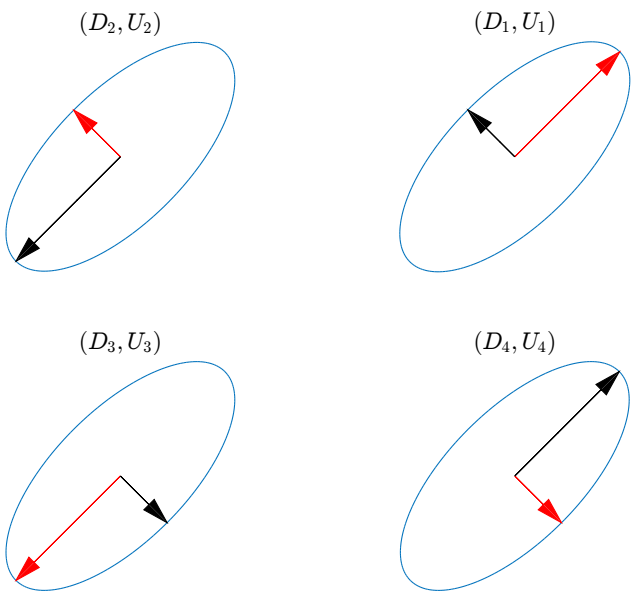

Figure 5: Depiction of the four eigen-decompositions of X.

2.8.4 Minimal Scaling-Rotation Framework

Define the eigen-composition map as $\mathcal{F} : Diag^+(p) \times SO(p) \mapsto Sym^+(p)$ as

$$\mathcal{F}(D, U) = UDU^T. \tag{2.12}$$

Given $X \in Sym^+(p)$, the set of all possible eigen-decompositions of X is

$$\mathcal{F}^{-1}(X) = \{(D, U) \in Diag^+(p) \times SO(p) : X = UDU^T\}.$$

For $X, Y \in Sym^+(p)$, one can define the distance between X and Y as the length of a minimal length scaling-rotation curve connecting members of $\mathcal{F}^{-1}(X)$ and $\mathcal{F}^{-1}(Y)$.

Definition 2.8.3. Given $X, Y \in Sym^+(p)$, the **minimal scaling-rotation** distance between X and Y is defined as

$$d_{MSR}(X, Y) := \inf_{\substack{(D_X, U_X) \in \mathcal{F}^{-1}(X), \\ (D_Y, U_Y) \in \mathcal{F}^{-1}(Y)}} d_{SR}((D_X, U_X), (D_Y, U_Y)), \tag{2.13}$$

where $d_{SR}(.,.)$ is the scaling-rotation distance function defined in Equation 2.11.

Remark 2.8.3. The function $d_{MSR}(.,.)$ is a semi-metric on $Sym^+(p)$ (the triangle inequality fails to hold over all of $Sym^+(p)$) and is a proper metric when restricted to the set of SPD matrices with all distinct eigenvalues.

Remark 2.8.4. The method for computing the minimal scaling-rotation between $X, Y \in Sym^+(p)$ depends on the eigenvaule multiplicity patterns of X and Y. Procedures for computing the minimal scaling-rotation distance for $p = 2$ and 3 were introduced in [28] and later improved upon in [22].

The minimal scaling-rotation distance between $X, Y \in Sym^+(p)$ measures the length of a minimal-length scaling-rotation curve on the product space $Diag^+(p) \times SO(p)$ connecting members of $\mathcal{F}^{-1}(X)$ and $\mathcal{F}^{-1}(Y)$. Intuitively, it can be thought of as the minimal amount of rotation and scaling required to deform the ellipsoid associated with X into the ellipsoid associated with Y. If $(D_X^*, U_X^*) \in \mathcal{F}^{-1}(X)$ and $(D_Y^*, U_Y^*) \in \mathcal{F}^{-1}(Y)$ are endpoints of a minimal length curve among all curves connecting eigen-decompositions of X and Y, we will refer to (D_X^*, U_X^*) and (D_Y^*, U_Y^*) as a **minimal pair** of eigen-decompositions for X and Y.

Definition 2.8.4. Let $(D_X, U_X), (D_Y, U_Y)$ be a minimal pair of eigen-decompositions for $X, Y \in Sym^+(p)$. A **minimal scaling-rotation curve** running from X to Y on $Sym^+(p)$ is of the form

$$\chi_{MSR}(t) = (Exp(Log(U_Y U_X^T)t)U_X)(Exp(Log(D_Y D_X^{-1})t)D_X)(Exp(Log(U_Y U_X^T)t)U_X)^T \tag{2.14}$$

for $t \in [0, 1]$.

As an illustration of how to compute the minimal scaling-rotation distance between two SPD matrices, let

$$X = \begin{pmatrix} 10 & 0 \\ 0 & 2 \end{pmatrix} \text{ and } Y = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 10 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

The four possible scaling-rotation curves connecting X and Y on $Sym^+(2)$ are displayed in Figure 6. Each row in the figure corresponds to a different scaling-rotation curve and can be read from left to right to visualize deformation from X, represented by the leftmost ellipse in each row, to Y, which is represented by the rightmost ellipse in each row. The heading

above each row displays the length of its corresponding scaling-rotation curve computed via Equation 2.11 with $k = 1$. It is no surprise that the deformation paths in the first two rows have the shortest scaling-rotation distances since those two paths are pure rotations from X to Y, while the bottom two rows appear to follow the same rotation patterns plus scaling. The deformation paths in the first two rows represent the two possible ways to rotate the principal axes of X to align with those of Y: clockwise rotation by 135 degrees (row 1) or counterclockwise rotation by 45 degrees (row 2). Note that since the scaling-rotation curves in the first two rows are pure rotations, their scaling-rotation lengths equal the total angle of rotation from X to Y along each path, and we have that the curve in row two is the minimal scaling-rotation curve connecting X and Y, with minimal scaling-rotation distance equal to 45 degrees or $\pi/4$ radians.

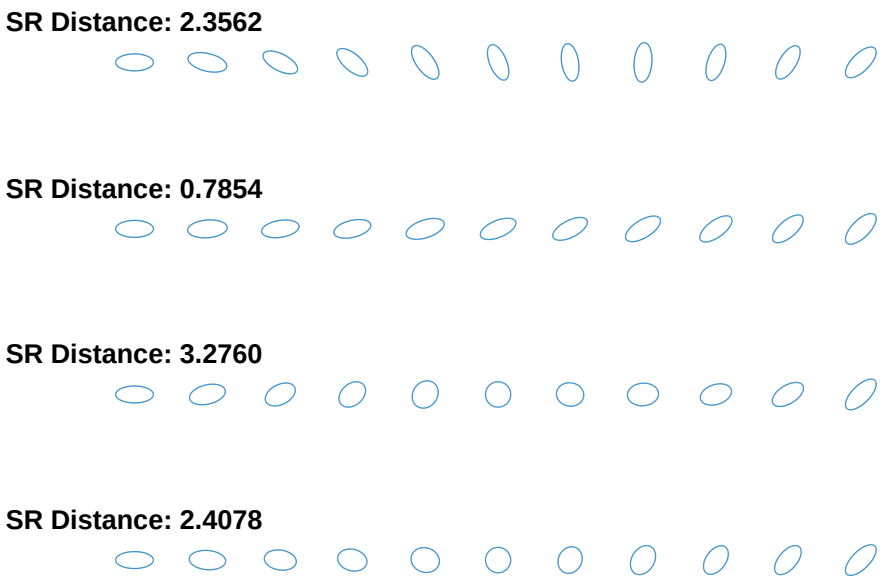

Figure 6: The four possible scaling-rotation curves connecting X and Y.

Note that in the example illustrated in Figure 6, there were four distinct scaling-rotation curves of four distinct lengths connecting X and Y and hence, only one minimal scaling-rotation curve running from X to Y. In the supplementary section of [22], the authors provide examples of pairs of SPD matrices from $Sym^+(2)$ and $Sym^+(3)$ that are connected by more than one minimal scaling-rotation curve. A pair of SPD matrices from $Sym^+(2)$

can be connected by one, two, or three distinct minimal scaling-rotation curves, and explicit conditions for each case are provided in the aforementioned supplementary section. While explicit conditions for each possible number of distinct minimal scaling-rotation curves connecting a pair of SPD matrices from $Sym^+(3)$ is not yet known (this would be an extremely tedious combinatorial exercise), the authors provide their most recent "worst case" of non-unique minimal scaling-rotation curves with an example of a pair of SPD matrices from $Sym^+(3)$ that are connected by nine distinct minimal length scaling-rotation curves. While the non-uniqueness of minimal scaling rotation curves connecting two SPD matrices can occur, it is virtually impossible to be observed in real data since the non-uniqueness requires equality of functions depending on the eigenvalues and eigenvectors of the two matrices.

We conclude this section with an informal description of the geometric stucture of $Sym^+(p)$ when equipped the minimal scaling-rotation distance function. For a thorough and rigorous description of the geometry of $Sym^+(p)$ under the minimal scaling-rotation framework, we refer readers to [22]. Eigen-decomposition partitions $Sym^+(p)$ according to eigenvalue multiplicity patterns. For example, using the notation from [22], $Sym^+(3) = \mathcal{S}^{top} \cup \mathcal{S}^{mid} \cup \mathcal{S}^{bot}$ where $\mathcal{S}^{top}$ denotes the set of 3×3 SPD matrices having 3 distinct eigenvalues, $\mathcal{S}^{mid}$ denotes the set of 3×3 SPD matrices with only two distinct eigenvalues, and $\mathcal{S}^{bot}$ denotes the set of SPD matrices having only one distinct eigenvalue with multiplicity 3. The spaces $\mathcal{S}^{top}, \mathcal{S}^{mid}$, and $\mathcal{S}^{bot}$ are disjoint manifolds of different dimensions, which characterizes $Sym^+(3)$ as a stratified space. For any $X \in Sym^+(3)$, the structure of $\mathcal{F}^{-1}(X)$ will vary depending on whether X belongs to $\mathcal{S}^{top}, \mathcal{S}^{mid}$, or $\mathcal{S}^{bot}$, which explains why the computation of $d_{MSR}(X, Y)$ depends on the respective eigenvalue multiplicity patterns of X and Y.

This informal description of the minimal scaling-rotation geometric framework can be adapted for general $p \neq 3$ by noting that $Sym^+(2) = \mathcal{S}^{top} \cup \mathcal{S}^{bot}$ and $Sym^+(p) = \mathcal{S}^{top} \cup \mathcal{S}^{mid} \cup \mathcal{S}^{bot}$ for $p > 3$, where $\mathcal{S}^{mid}$ is the union of subsets of $Sym^+(p)$ corresponding to all possible eigenvalue multiplicity patterns having at least two distinct eigenvalues and at least one repeated eigenvalue. It is important to note that while the cross-product space $Diag^+(p) \times SO(p)$ is a Riemannian manifold when equipped with the scaling rotation distance from Equation 2.11, $Sym^+(p)$ is not a Riemannian manifold under the minimal scaling-rotation framework [22].

3.0 SCALING-ROTATION ESTIMATION FOR SYM-PLUS(P)

3.1 MINIMAL SCALING-ROTATION SAMPLE MEAN ESTIMATION

In this chapter we consider the setting in which we have a random sample of SPD matrices and would like to compute a sample-based location estimator. Location estimation is an important first step in the development of many statistical techniques including two sample hypothesis testing ([45]) for comparing average brain scans from two groups of interest, principal geodesic analysis ([18]) for visualizing major modes of variation in a sample of SPD matrices, and weighted mean estimation, which has useful applications in diffusion tensor processing, including fiber tracking, smoothing, and interpolation ([4],[9]).

One of the challenges of developing methods for analyzing SPD-valued data is that the positive-definite constraint precludes $\text{Sym}^+(p)$ from being a vector space. This can be easily seen when $p = 2$, as plotting the free coordinates (two diagonal elements and upper off-diagonal element) of all 2×2 SPD matrices in $\mathbb{R}^3$ fills the interior of a convex cone. Hence, conventional estimation or inferential techniques developed for data that varies freely over Euclidean space may not be appropriate for the statistical analysis of SPD matrices. With this in mind, many location estimation frameworks that account for the geometry of $\text{Sym}^+(p)$ have been developed in recent years, including the Log-Euclidean framework ([2]), Affine-Invariant framework ([18], [40]), and Procrustes Size and Shape framework ([15]). Given a sample of SPD matrices, most of these estimation methods amount to transforming the SPD-valued observations, averaging in the space of the transformed observations, and then mapping the mean of the transformed data into $\text{Sym}^+(p)$. For example, the Log-Euclidean method maps each observation into $\text{Sym}(p)$, the space of $p \times p$ symmetric matrices, via the matrix logarithm, computes the sample mean of the transformed observations, and then

maps that mean into $\text{Sym}^+(p)$ via the matrix exponential function, while the Procrustes Size and Shape method begins with computing a matrix square root for each observation, averages the matrix square roots in a way that accounts for the non-uniqueness of the matrix square root, and then maps the final average $\hat{L}$ to $\text{Sym}^+(p)$ as $\hat{\Sigma} = \hat{L}\hat{L}^T$.

While these geometric frameworks account for the curvature of $\text{Sym}^+(p)$, it is not clear which of the Log-Euclidean, Affine-Invariant, or Procrustes Size-and-Shape is most "natural" for describing deformations of SPD matrices. Motivated by the analysis of DTI data, a setting in which observations are ellipsoids in $\mathbb{R}^3$, Jung et al. ([28]) developed the scaling-rotation framework for $\text{Sym}^+(p)$. Under this framework, the distance between SPD matrices X and Y is defined as the minimal amount of rotation of axes and scaling of axis lengths necessary to defom the ellipsoid associated with X into the ellipsoid associated with Y. The authors found that the scaling-rotation framework yields interpolation curves that have desirable properties, including constant rate of rotation and log-linear scaling of eigenvalues, and found it to be the only geometric framework (compared to Euclidean, Log-Euclidean, and Affine-Invariant) to produce both pure scaling interpolation curves and pure rotation curves when the endpoints differed by pure scaling or pure rotation.

Given that no statistical methods yet exist for the scaling-rotation framework, the goal of this chapter is to establish location estimation methods using the scaling-rotation framework as a foundation for future methods that will inherit the interpretability of the framework. In this chapter, we provide computational procedures for computing a sample scaling-rotation mean set and a weighted sample scaling-rotation mean set, and list conditions that guarantee uniqueness, strong consistency, and a type of Central Limit Theorem for sample scaling-rotation means. We conclude the chapter with some real data applications of our scaling-rotation estimation framework to multivariate tensor-based morphometry and DTI data processing.

3.1.1 Fréchet Mean Estimation

An approach often used for developing location estimators for non-Euclidean metric spaces is Fréchet mean estimation [19], in which estimators are derived as minimizers of a metric-

dependent sample mean-squared error. In this subsection, we provide definitions of population and sample Fréchet means, as well as references to examples of well-known Fréchet mean estimators for $\text{Sym}^+(p)$.

Definition 3.1.1. Let M be a metric space with metric ρ and suppose that $X, X_1, \ldots, X_n$ are i.i.d. M-valued random variables with probability measure μ. The *population Fréchet mean set* is

$$\underset{C \in M}{\text{argmin}} \int_M \rho^2(X, C) d\mu(X).$$

The *sample Fréchet mean set* is

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

The Fréchet mean estimation framework can viewed as an extension of location estimation on the real line to location estimation on metric spaces since the population Fréchet mean and sample Fréchet mean equal the expected value of a random variable and arithmetic mean of a random sample, respectively, when $M = \mathbb{R}$ and $\rho(x, y) = |x - y|$.

Examples of location estimators that have been developed for $\text{Sym}^+(p)$ using the sample Fréchet mean estimation framework include the Log-Euclidean mean ([2]), Affine-Invariant mean ([18], [40]), and Procrustes Size and Shape mean ([15]).

3.1.2 Sample Scaling-Rotation Mean

In this section, we define the sample minimal scaling-rotation mean set as the sample Fréchet mean set of a sample of SPD matrices under the scaling-rotation framework, introduce a related estimator called the sample partial minimal scaling-rotation mean set which often coincides with the sample minimal scaling-rotation mean set, and then provide a procedure for computing a candidate member of the sample partial minimal scaling-rotation mean set.

Definition 3.1.2. Let $S_1, \ldots, S_n$ be a random sample of SPD matrices. The *sample minimal scaling-rotation mean set* is

$$\underset{\Sigma \in Sym^+(p)}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} d^2_{MSR}(S_i, \Sigma; K).$$

Solving for the sample minimal scaling-rotation mean set will be challenging in practice. Suppose that $(D_i^*, U_i^*) \in \mathcal{F}^{-1}(S_i)$ and $(D_\Sigma^{(i)}, U_\Sigma^{(i)}) \in \mathcal{F}^{-1}(\Sigma)$ form minimal pairs for $i = 1, \ldots, n$. Then $\frac{1}{n}\sum_{i=1}^n d_{MSR}^2(S_i, \Sigma) = \frac{1}{n}\sum_{i=1}^n d_{SR}^2((D_i^*, U_i^*), (D_\Sigma^{(i)}, U_\Sigma^{(i)}))$, and there is no guarantee that $(D_\Sigma^{(i)}, U_\Sigma^{(i)}) = (D_\Sigma^{(j)}, U_\Sigma^{(j)})$ for any pair i, j, which complicates optimization over $M(p)$. To address this issue, note that the minimal scaling-rotation distance can be simplified as

$$d_{MSR}(S, \Sigma) = \inf_{(D_S, U_S) \in \mathcal{F}^{-1}(S)} d_{SR}((D_S, U_S), (D, U)), \tag{3.1}$$

where (D, U) is any eigen-decomposition of Σ when S and Σ both have no repeated eigenvalues. Motivated by this simplification, we define the following measure of distance between an SPD matrix and a given eigen-decomposition:

Definition 3.1.3. The *partial minimal scaling-rotation distance* is $d_{\mathcal{PMSR}} : \mathrm{Sym}^+(p) \times M(p) \mapsto [0, \infty)$ given by

$$d_{\mathcal{PMSR}}(S, (D, U)) = \inf_{(D_S, U_S) \in \mathcal{F}^{-1}(S)} d_{SR}((D_S, U_S), (D, U)).$$

Definition 3.1.4. Let $S_1, \ldots, S_n$ be a random sample of SPD matrices. The *sample partial minimal scaling-rotation mean set* is

$$\operatorname*{argmin}_{(D,U) \in M(p)} \frac{1}{n} \sum_{i=1}^n d_{\mathcal{PMSR}}^2(S_i, (D, U)).$$

Remark 3.1.1. Note that the sample partial minimal scaling-rotation mean set belongs to the product space $M(p) = Diag^+(p) \times SO(p)$, while the sample minimal scaling-rotation mean set belongs to $Sym^+(p)$. If one prefers to define the sample partial minimal scaling-rotation mean set on $Sym^+(p)$, each memeber of the set can be mapped to $\mathrm{Sym}^+(p)$ via eigen-composition.

The partial scaling-rotation sample mean set should, in practice, provide a good approximation of the eigen-decompositions of the scaling-rotation sample mean set since those sets are equivalent, by (3.1), when each observation has no repeated eigenvalues and the parameter space is restricted to the set of SPD matrices which have no repeated eigenvalues. Expecting all observations to have no repeated eigenvalues in practice is not unreasonable, as this will be the case with probability 1 if the sample arises from a continuous distribution.

Let $Q(D,U) = \frac{1}{n}\sum_{i=1}^{n} d^2_{\mathcal{PMSR}}(S_i,(D,U))$. We propose the following procedure, which is similar to the Generalized Procrustes Algorithm ([21]), for approximating a member of the sample partial scaling-rotation mean set:

1. Set tolerance $\varepsilon > 0$ and pick initial guess $(\hat{D}^{(0)}, \hat{U}^{(0)})$.
2. For $k \geq 0$ and $i = 1, \ldots, n$, find $(D_i^{(k)}, U_i^{(k)}) \in \mathcal{F}^{-1}(S_i)$ that has the smallest scaling-rotation distance from $(\hat{D}^{(k)}, \hat{U}^{(k)})$.
3. Compute
$$(\hat{D}^{(k+1)}, \hat{U}^{(k+1)}) \in \underset{(D,U)\in M(p)}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} d^2_{SR}((D_i^{(k)}, U_i^{(k)}),(D,U))$$
4. If $|Q(\hat{D}^{(k+1)}, \hat{U}^{(k+1)}) - Q(\hat{D}^{(k)}, \hat{U}^{(k)})| > \varepsilon$, repeat steps 2 and 3.
   Otherwise, set $(\hat{D}_{\mathcal{PMSR}}, \hat{U}_{\mathcal{PMSR}}) = (\hat{D}^{(k+1)}, \hat{U}^{(k+1)})$.

Remark 3.1.2. The above procedure will always terminate since $Q(D,U) \geq 0$ for any $(D,U) \in M(p)$ and $Q(\hat{D}^{(k)}, \hat{U}^{(k)}) \geq Q(\hat{D}^{(k+1)}, \hat{U}^{(k+1)})$ for any $k \geq 0$.

When observation S_i has no repeated eigenvalues, performing Step 2 will simply require searching over the $2^{(p-1)}p!$ distinct eigen-decompositions (cf. Theorem 3.3 of [28]) of S_i to find one that attains the minimal scaling-rotation distance from $(\hat{D}^{(k)}, \hat{U}^{(k)})$. Solving for the minimizing eigen-decomposition of S_i is also easy when S_i is a scaled identity matrix, since $S_i = cI_p = U(cI_p)U^T$ for any $U \in SO(p)$ implies that $(cI_p, \hat{U}^{(k)})$ will be the eigen-decomposition of S_i with minimal geodesic distance from $(\hat{D}^{(k)}, \hat{U}^{(k)})$. Determining the minimizing eigen-decomposition of S_i when $p = 3$ and S_i has two unique eigenvalues can be done by using a similar computational method as in part (iii) of Theorem 4.3 from [28]. For $p > 3$ and S_i having more than 1 but fewer than p unique eigenvalues, determining the minimal eigen-decomposition of S_i for step 2 is an open problem. Since the number of eigen-decompositions ($2^{(p-1)}p!$) of a SPD matrix which has no repeated eigenvaues grows rapidly with p, we intend to use this procedure for $p = 2, 3$. Performing Step 2 for larger p may require numerical procedures that can circumvent having to search over all $2^{(p-1)}p!$ eigen-decompositions for each observation.

The optimization problem over $M(p)$ from Step 3 can be divided into separate minimiza-

tion problems over $Diag^+(p)$ and $SO(p)$:

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

$\hat{D}^{(k+1)}$ has the closed-form solution $\hat{D}^{(k+1)} = Exp(\frac{1}{n}\sum_{i=1}^{n} Log(D_i^{(k)}))$. The solution for $\hat{U}^{(k+1)}$ must be approximated via numerical procedures, and it shown in [34] that when the rotation matrices $U_1^{(k)}, \ldots, U_n^{(k)}$ lie within a geodesic ball of radius $\frac{\pi}{2}$, there is a unique solution for $\hat{U}^{(k+1)}$ which can be approximated by a globally convergent gradient descent algorithm.

3.1.3 Weighted Scaling-Rotation Mean

Often it is useful to compute a weighted average of SPD matrices, especially in the processing of diffusion tensor data via fiber tracking, interpolation, or smoothing. The procedure for estimating the sample partial minimal scaling-rotation mean can easily be adapted to compute a weighted sample partial minimal scaling-rotation mean:

Definition 3.1.5. The *weighted partial minimal scaling-rotation sample mean set* is

$$\underset{(D,U) \in M(p)}{\operatorname{argmin}} \sum_{i=1}^{n} w_i d^2_{\mathcal{PMSR}}(S_i, (D, U)).$$

where positive weights $w_1, \ldots, w_n$ satisfy $\sum_{i=1}^{n} w_i = 1$.

The procedure for finding a candidate member of the sample partial minimal scaling-rotation mean set can be modified to compute a candidate member of the weighted sample partial minimal scaling-rotation mean set, which involves changing step 3 so that

$$\hat{D}^{(k+1)} = \underset{D \in Diag^+(p)}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \|Log(D_i^{(k)}) - Log(D)\|_F^2$$
$$\hat{U}^{(k+1)} \in \underset{U \in SO(p)}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \|Log(U_i^{(k)} U^{-1})\|_F^2.$$

$\hat{D}^{(k+1)}$ has the closed form solution $\hat{D}^{(k+1)} = Exp(\sum_{i=1}^{n} w_i Log(D_i^{(k)}))$ and $\hat{U}^{(k+1)}$ can be computed by modifying the gradient descent method of [34] to solve for $U \in SO(p)$ satisfying $\sum_{i=1}^{n} w_i Log(U_i^{(k)} U^{-1}) = \mathbf{0}$.

3.2 THEORETICAL PROPERTIES OF THE SCALING-ROTATION SAMPLE MEAN

In this section we assume that we have an i.i.d. random sample $S_1, \ldots, S_n$ of SPD matrices which have no repeated eigenvalues. The notation S_p^{top} will denote the set of $p \times p$ SPD matrices which have no repeated eigenvalues, $E^{(\mathcal{PMSR})}$ will denote the set of population partial minimal scaling-rotation means, and $E_n^{(\mathcal{PMSR})}$ will denote the set of sample partial minimal scaling-rotation sample means.

3.2.1 Even Signed Permutations

To set up the main theoretical properties of the sample partial scaling-rotation mean, in this subsection we formally describe the set of eigen-decompositions of an SPD matrix from S_p^{top} in relation to a special finite subgroup of $SO(p)$, which we define as the set of even signed permutation matrices. At the end of this section, we define a descriptor of the set of even signed permutation matrices which, when scaled, provides a lower bound on the distance between distinct eigen-decompositions. This descriptor will be used in conditions for uniqueness of the sample partial scaling-rotation mean and a Central Limit Theorem (CLT) in the subsequent subsections.

Recall from Theorem 2.8.3 that an SPD matrix from S_p^{top} has $2^{p-1}p!$ distinct eigen-decompositions, which are related via permutations of eigenvalues and simultaneous sign-changes and permutations of eigenvectors. More precisely, given $(D, U) \in \mathcal{F}^{-1}(X)$, any other eigen-decomposition $(D', U') \in \mathcal{F}^{-1}(X)$ has the form $(D', U') = (P_\pi^+ D P_\pi^{+T}, U I_\sigma P_\pi^{+T})$, where I_σ is a positive-determinant sign change matrix and P_π^+ is a modified permutation matrix.

To simplify notation when describing the set of eigen-decompositions of an SPD matrix from S_p^{top}, we define the set of even signed permutation matrices:

Definition 3.2.1. The set of $p \times p$ *even signed permutation matrices* is

$$\mathcal{G}(p) = \{W \in SO(p) : W = P_\pi^+ I_\sigma, \text{ where } P_\pi^+ \in \mathcal{P}^+(p), I_\sigma \in \mathcal{A}(p), \}$$

where $\mathcal{P}^+(p)$ denotes the set of $p \times p$ modified permutation matrices and $\mathcal{A}(p)$ denotes the set of $p \times p$ positive-definite sign-change matrices.

Remark 3.2.1. In [23], the authors use the notation $\tilde{S}_p^+$ to denote the set of $p \times p$ signed permutation matrices.

$\mathcal{G}(p)$ is a group under multiplication and is a subgroup of $SO(p)$ of order $2^{p-1}p!$. We define the action of $\mathcal{G}(p)$ on $M(p)$ via the map $\eta : \mathcal{G}(p) \times M(p) \mapsto M(p)$ with

$$\eta(G, (D, U)) = (GDG^T, UG^T).$$

If (D_X, U_X) is an eigen-decomposition of an SPD matrix from S_p^{top}, then the orbit of (D_X, U_X), which we denote as $\eta(\mathcal{G}(p), (D_X, U_X)) = \{\eta(G, (D_X, U_X)) : G \in \mathcal{G}(p)\}$, is the set of eigen-decompositions of X. This orbit structure is also useful for describing the set of sample partial minimal scaling-rotation means when each observation in the sample comes from S_p^{top} since it can be shown that

$$\frac{1}{n}\sum_{i=1}^{n} d^2_{\mathcal{PMSR}}(S_i, (D, U)) = \frac{1}{n}\sum_{i=1}^{n} d^2_{\mathcal{PMSR}}(S_i, \eta(G, (D, U))) \tag{3.2}$$

for any $G \in \mathcal{G}(p)$ when $S_1, \ldots, S_n \in S_p^{top}$. It follows from (3.2) that for any $(D^*, U^*) \in E_n^{(\mathcal{PMSR})}$, its orbit $\eta(\mathcal{G}(p), (D^*, U^*))$ also belongs to $E_n^{(\mathcal{PMSR})}$. Thus, $E_n^{(\mathcal{PMSR})}$ will contain at least $2^{(p-1)}p!$ elements whenever a solution exists, and in the case when $E_n^{(PMSR)}$ only contains $2^{(p-1)}p!$ elements which all belong to the same orbit, we will say that the sample partial minimal scaling-rotation mean is unique up to the action of $\mathcal{G}(p)$.

The following result concerning the distance between eigen-decompositions of an SPD matrix from S_p^{top} will be useful in subsequent subsections:

Lemma 3.2.1. *For any $X \in S_p^{top}$, any two distinct eigen-decompositions (D_X, U_X) and (D'_X, U'_X) of X satisfy*

$$d_M((D_X, U_X), (D'_X, U'_X)) \geq \sqrt{K}\beta_{\mathcal{G}(p)},$$

where K is the rotation distance multiplier from Equation 2.11 and $\beta_{\mathcal{G}(p)}$ is defined as

$$\beta_{\mathcal{G}(p)} = \min_{G \in \mathcal{G}(p)\setminus\{I_p\}} \frac{1}{\sqrt{2}}\|Log(G)\|_F.$$

For any $p \geq 2$, it is true that $\beta_{\mathcal{G}(p)} \leq \pi/2$.

Proof. Given $X \in S_p^{top}$, let (D_X, U_X) and (D'_X, U'_X) be two distinct eigen-decompositions of X. From Theorem 2.8.3, there is an even signed-permutation matrix G such that $(D'_X, U'_X) = (GDG^T, UG^T)$. Since (D_X, U_X) and (D'_X, U'_X) are distinct, $G \neq I_p$. Then we have that

$$\begin{aligned}
&d_{SR}((D_X, U_X), (D'_X, U'_X)) \\
&= (d^2_{\mathcal{D}^+(p)}(D_X, GD_XG^T) + Kd^2_{SO(p)}(U_X, U_XG^T))^{1/2} \\
&\geq \sqrt{K} d_{SO(p)}(U_X, U_XG^T) \\
&= \sqrt{K} d_{SO(p)}(I_p, G^T) \\
&\geq \sqrt{K}\beta_{\mathcal{G}(p)}.
\end{aligned}$$

We next show that $\beta_{\mathcal{G}(p)} \leq \pi/2$ for any $p \geq 2$. The set $\mathcal{G}(p)$ contains the block diagonal signed permutation matrix

$$B = \begin{pmatrix} I_{p-2} & 0 \\ 0 & R(\frac{\pi}{2}) \end{pmatrix},$$

where

$$R\left(\frac{\pi}{2}\right) = \begin{pmatrix} \cos(\frac{\pi}{2}) & -\sin(\frac{\pi}{2}) \\ \sin(\frac{\pi}{2}) & \cos(\frac{\pi}{2}) \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

It can be shown that

$$Log(B) = \begin{pmatrix} 0 & 0 \\ 0 & Log(R(\frac{\pi}{2})) \end{pmatrix},$$

where

$$Log\left(R\left(\frac{\pi}{2}\right)\right) = \begin{pmatrix} 0 & -\frac{\pi}{2} \\ \frac{\pi}{2} & 0 \end{pmatrix}.$$

Then we have that $d_{SO(p)}(I_p, B) = \pi/2$, implying that $\beta_{\mathcal{G}(p)} \leq \pi/2$. □

3.2.2 Uniqueness of the Partial Minimal Scaling-Rotation Sample Mean

Much work has been done on the question of uniqueness of the sample Fréchet mean of Riemannian manifold-valued observations, coming to the conclusion that the the sample Fréchet mean will be unique as long as the observations lie within a geodesic ball of a certain radius (see, for example, [3]). While $\text{Sym}^+(p)$ is not a Riemannian manifold when equipped

with $d_{\mathcal{SR}}$, we can obtain a similar result for a kind of uniqueness of the sample partial scaling-rotation mean. Recall that if $(D^*, U^*) \in E_n^{(\mathcal{PMSR})}$, then $\eta(\mathcal{G}(p), (D^*, U^*)) \subset E_n^{(\mathcal{PMSR})}$ by (3.2). This means that, at best, a sample partial minimal scaling-rotation mean is unique up to the action of $\mathcal{G}(p)$, which occurs when there are $2^{(p-1)}p!$ sample partial scaling-rotation means belonging to the same orbit.

In the subsequent theorem, we show that if the sample observations are sufficiently concentrated around the initial guess, then the sample mean estimation procedure from Section 3.1.2 will yield a solution that is unique up to the action of $\mathcal{G}(p)$, and will converge after two iterations.

Theorem 3.2.1. *Assume $S_1, \ldots, S_n \in S_p^{top}$. If*

$$d_{\mathcal{MSR}}(S_i, S_j) < \frac{\beta_{\mathcal{G}(p)}\sqrt{K}}{2(1+\sqrt{2})} \tag{3.3}$$

for $i, j = 1, \ldots n$, then choosing an eigen-decomposition of any observation from the sample as the initial solution will lead to a unique solution, up to the action of $\mathcal{G}(p)$, and the estimation procedure will terminate after two iterations, with $(\hat{D}^{(1)}, \hat{U}^{(1)}) = (\hat{D}^{(2)}, \hat{U}^{(2)}) = (\hat{D}_{\mathcal{PMSR}}, \hat{U}_{\mathcal{PMSR}})$.

Proof. Without loss of generality, set the initial guess $(\hat{D}_0, \hat{U}_0)$ to be any eigen-decomposition of S_1. For $i = 1, \ldots, n$ choose $(D_i^{(0)}, U_i^{(0)}) \in \mathcal{F}^{-1}(S_i)$ that forms a minimal pair with $(\hat{D}_0, \hat{U}_0)$, and then compute the scaling-rotation sample mean

$$(\hat{D}^{(1)}, \hat{U}^{(1)}) \in \underset{(D,U)\in M(p)}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} d_{SR}^2((D_i^{(0)}, U_i^{(0)}), (D, U)).$$

Let $C = \beta_{\mathcal{G}(p)}\sqrt{K}/(2(1+\sqrt{2}))$. From the definition of the scaling-rotation distance function, it is easy to verify that $d_{MSR}(S_i, S_1) < C$ for $i = 1, \ldots, n$ implies that

$$d_{\mathcal{D}^+(p)}(D_i^{(0)}, \hat{D}^{(0)}) < C \text{ and } \sqrt{K} d_{SO(p)}(U_i^{(0)}, \hat{U}^{(0)}) < C$$

for $i = 1, \ldots, n$. The eigenvalue matrix mean of the aligned eigenvalue matrices has the closed form expression $\hat{D}^{(1)} = \operatorname{Exp}(\frac{1}{n}\sum_{i=1}^{n} \operatorname{Log}(D_i^{(0)}))$. While there is no closed-form expression for the eigenvector matrix mean, there is a unique solution for $\hat{U}^{(1)}$ since $U_1^{(0)}, \ldots, U_n^{(0)}$ lie within a geodesic ball of radius less than $\pi/2$ centered at $\hat{U}^{(0)}$ (cf. Theorem 5 of [34]).

We first establish that the eigenvalue matrix mean $\hat{D}^{(1)}$ also lies within a geodesic ball of radius C centered at $\hat{D}^{(0)}$. Given $d_{\mathcal{D}^+(p)}(D_i^{(0)}, \hat{D}^{(0)}) < C$ for $i = 1, \ldots, n$, it follows that

$$\begin{aligned}
&d^2_{\mathcal{D}^+(p)}(\hat{D}^{(1)}, \hat{D}^{(0)}) \\
&= \left\| \frac{1}{n} \sum_{i=1}^{n} [\mathrm{Log}(\hat{D}_i^{(0)}) - \mathrm{Log}(\hat{D}^{(0)})] \right\|_F^2 \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} tr\left([\mathrm{Log}(D_i^{(0)}) - \mathrm{Log}(\hat{D}^{(0)})][\mathrm{Log}(D_j^{(0)}) - \mathrm{Log}(\hat{D}^{(0)})] \right) \\
&\leq \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{\mathcal{D}^+(p)}(D_i^{(0)}, \hat{D}^{(0)}) d_{\mathcal{D}^+(p)}(D_j^{(0)}, \hat{D}^{(0)}) \\
&< C^2.
\end{aligned}$$

The penultimate inequality follows from an application of the Cauchy-Schwarz inequality.

We next show that $U_1^{(0)}, \ldots, U_n^{(0)}$ lying with a geodesic ball of radius less than $C/\sqrt{K}$ centered at $\hat{U}^{(0)}$ implies that the eigenvector matrix mean $\hat{U}^{(1)}$ also has distance less than $C/\sqrt{K}$ from $\hat{U}^{(0)}$ by application of Theorem 2.1 from [3]. We need to show that $C/\sqrt{K} \leq \min\{inj(SO(p)), \pi/\sqrt{\triangle}\}/2$, where $inj(SO(p))$ denotes the injectivity radius of $SO(p)$ and $\triangle$ is an upper bound on the sectional curvatures of $SO(p)$. The Riemannian manifold $SO(p)$ equipped with metric $d_{SO(p)}$ has an exponential map with injectivity radius bounded below by π and sectional curvatures belonging to the closed interval $[0, 1/4]$ (Section 5 of [34]), which imply that $\min\{inj(SO(p)), \pi/\sqrt{\triangle}\}/2 \geq \pi/2$. Since $C/\sqrt{K} < \beta_{\mathcal{G}(p)} \leq \pi/2$, we have our desired bound for the radius of the geodesic ball containing $U_1^{(0)}, \ldots, U_n^{(0)}$, and it follows that $\hat{U}^{(1)}$ also has distance less than $C/\sqrt{K}$ from $\hat{U}^{(0)}$.

We continue the procedure for computing a candidate for sample partial minimal scaling-rotation mean by choosing $(D_i^{(1)}, U_i^{(1)}) \in \mathcal{F}^{-1}(S_i)$ that forms a minimal pair with $(\hat{D}^{(1)}, \hat{U}^{(1)})$ for $i = 1, \ldots, n$. For any observation S_i, we can bound $d_{SR}((D_i^{(0)}, U_i^{(0)}), (D_i^{(1)}, U_i^{(1)}))$ as follows

$$\begin{aligned}
&d_{SR}((D_i^{(0)}, U_i^{(0)}), (D_i^{(1)}, U_i^{(1)})) \\
&\leq d_{SR}((D_i^{(0)}, U_i^{(0)}), (\hat{D}^{(1)}, \hat{U}^{(1)})) + d_{SR}((D_i^{(1)}, U_i^{(1)}), (\hat{D}^{(1)}, \hat{U}^{(1)})) \\
&\leq 2 d_{SR}((D_i^{(0)}, U_i^{(0)}), (\hat{D}^{(1)}, \hat{U}^{(1)})) \\
&\leq 2 d_{SR}((D_i^{(0)}, U_i^{(0)}), (\hat{D}^{(0)}, \hat{U}^{(0)})) + 2 d_{SR}((\hat{D}^{(0)}, \hat{U}^{(0)}), (\hat{D}^{(1)}, \hat{U}^{(1)})) \\
&< 2(1 + \sqrt{2})C
\end{aligned}$$

Most of the inequalities above are applications of the Triangle Inequality, and the last inequality follows from $d_{MSR}(S_i, S_1) < C$ for $i = 1, \ldots, n$ and that $d_{\mathcal{D}^+(p)}(\hat{D}^{(1)}, \hat{D}^{(0)}) < C$ and $d_{SO(p)}(\hat{U}^{(1)}, \hat{U}^{(0)}) < C/\sqrt{K}$ imply that $d_{SR}((\hat{D}^{(0)}, \hat{U}^{(0)}), (\hat{D}^{(1)}, \hat{U}^{(1)})) < \sqrt{2}C$. Plugging in the value for C, we get $d_{SR}((D_i^{(0)}, U_i^{(0)}), (D_i^{(1)}, U_i^{(1)})) < \beta_{\mathcal{G}(p)}\sqrt{K}$. By Lemma 3.2.1, we have that $(D_i^{(0)}, U_i^{(0)}) = (D_i^{(1)}, U_i^{(1)})$ for $i = 1, \ldots, n$, which implies that $(\hat{D}^{(1)}, \hat{U}^{(1)})$ will equal $(\hat{D}^{(2)}, \hat{U}^{(2)})$, and the estimation procedure will terminate after two iterations. The final solution for the sample partial minimal scaling-rotation mean will be $(\hat{D}_{\mathcal{PMSR}}, \hat{U}_{\mathcal{PMSR}}) = (\hat{D}^{(1)}, \hat{U}^{(1)}) = (\hat{D}^{(2)}, \hat{U}^{(2)})$. □

Remark 3.2.2. The initial guess need not be an eigen-decomposition of a sample observation, as the conclusion of Theorem 3.2.1 will also follow from picking an initial guess $(\hat{D}^{(0)}, \hat{U}^{(0)}) \in M(p)$ satisfying

$$d_{\mathcal{PMSR}}(S_i, (\hat{D}^{(0)}, \hat{U}^{(0)})) < \frac{\beta_{\mathcal{G}(p)}\sqrt{K}}{2(1+\sqrt{2})}$$

for $i = 1, \ldots, n$.

Remark 3.2.3. If $(\hat{D}_1^{(0)}, \hat{U}_1^{(0)}), (\hat{D}_2^{(0)}, \hat{U}_2^{(0)}) \in M(p)$ satisfy

$$d_{\mathcal{PMSR}}(S_i, (\hat{D}_j^{(0)}, \hat{U}_j^{(0)})) < \frac{\beta_{\mathcal{G}(p)}\sqrt{K}}{2(1+\sqrt{2})}$$

for $i = 1, \ldots, n$ and $j = 1, 2$, then choosing $(\hat{D}_1^{(0)}, \hat{U}_1^{(0)})$ or $(\hat{D}_2^{(0)}, \hat{U}_2^{(0)})$ as the initial guess for the sample partial scaling-rotation mean will lead to the same final solution for the sample partial scaling-rotation mean set.

Given an initial guess $(\hat{D}^{(0)}, \hat{U}^{(0)})$ for a member of the sample partial minimal scaling-rotation mean set, let $(D_i^{(0)}, U_i^{(0)})$ denote the eigen-decomposition of S_i that has minimal scaling-rotation distance from the initial guess. Note that the concentration condition

$$d_{\mathcal{PMSR}}(S_i, (\hat{D}^{(0)}, \hat{U}^{(0)})) < \frac{\beta_{\mathcal{G}(p)}\sqrt{K}}{2(1+\sqrt{2})} \tag{3.4}$$

for $i = 1, \ldots, n$ of Theorem 3.2.1 implies that

$$d_{\mathcal{D}^+(p)}(D_i^{(0)}, \hat{D}^{(0)}) < \frac{\beta_{\mathcal{G}(p)}\sqrt{K}}{2(1+\sqrt{2})} \text{ and } d_{SO(p)}(U_i^{(0)}, \hat{U}^{(0)}) < \frac{\beta_{\mathcal{G}(p)}}{2(1+\sqrt{2})}$$

for $i = 1, \ldots, n$. In samples where Theorem 3.2.1 holds, K affects the amount of variability of the sample aligned eigenvalue matrices $D_1^{(0)}, \ldots, D_n^{(0)}$ about $\hat{D}^{(0)}$. When K is very large

(small), the sample aligned eigenvalue matrices will belong to a ball centered at $\hat{D}^{(0)}$ with a very large (small) radius. This makes intuitive sense because K controls the importance of rotation distance relative to scaling distance in the computation of the scaling-rotation distance: when K is much larger than 1, the minimal scaling-rotation curves connecting two SPD matrices will have minimal rotation of eigenvectors with little regards to the amount of scaling of eigenvalues, and vice versa when K is very small compared to 1.

3.2.3 Asymptotic Properties of the Scaling-Rotation Sample Mean

In this section we address two aspects of the asymptotic behavior of the sample partial minimal scaling-rotation mean: (i) strong consistency of the sample partial minimal scaling-rotation mean set to the population partial minimal scaling-rotation mean set and (ii) the limiting distribution of the sample partial minimal scaling-rotation mean as $n \to \infty$. Much work has been done to establish consistency and CLT-type results for sample Fréchet means on Riemannian manifolds and metric spaces ([6], [7], [5]), however, estimation of the partial minimal scaling-rotation mean does not fit into the context of estimation on Riemannian manifolds nor metric spaces since the sample space, S_p^{top}, and parameter space, $M(p)$ are different. With this in mind, we applied the asymptotic results for sample Fréchet means on general product spaces in [26] and [27] to partial scaling-rotation mean estimation to establish conditions for strong consistency and a Central Limit Theorem.

Lemma 3.2.2. *The following properties hold for $d_{\mathcal{PMSR}}$:*

1. *(Uniform Continuity) For every $\varepsilon > 0$ there is a $\delta(\varepsilon) > 0$ such that for every $X \in S_p^{top}$, $d_{SR}((D,U),(D',U')) < \delta$ implies $|d_{\mathcal{PMSR}}(X,(D,U);K) - d_{\mathcal{PMSR}}(X,(D',U');K)| < \varepsilon$.*
2. *(Continuity) $d_{\mathcal{PMSR}}$ is continuous on $S_p^{top} \times M(p)$.*
3. *(Coercivity) For any $X \in S_p^{top}$, $(D_0,U_0) \in M(p)$, and any sequence $(D_n,U_n) \in M(p)$ with $d_M((D_n,U_n),(D_0,U_0)) \to \infty$, it follows that $d_{\mathcal{PMSR}}(X,(D_n,U_n)) \to \infty$.*

Proof. 1. From Exercise 20 of Ch. 4 from [41], we know that if A is any non-empty subset of a metric space (M,d), then the function $f(y) = \inf_{x \in A} d(x,y)$ satisfies

$$|f(y) - f(y')| \leq d(y,y')$$

for any $y, y' \in M$.

From the fact that $(M(p), d_{SR})$ is a metric space and $\mathcal{F}^{-1}(X) \subset M(p)$ is non-empty for any $X \in Sym^+(p)$, and the definition of $d_{\mathcal{PMSR}}$, we have that

$$|d_{\mathcal{PMSR}}(X,(D,U)) - d_{\mathcal{PSR}}(X,(D',U'))| \leq d_M((D,U),(D',U'))$$

for any $(D,U),(D',U') \in M(p)$.

Fix $\varepsilon > 0$ and set $\delta = \varepsilon$. Then we have that $d_{SR}((D,U),(D',U')) < \delta$ implies that $|d_{\mathcal{PMSR}}(X,(D,U)) - d_{\mathcal{PMSR}}(X,(D',U'))| < \varepsilon$.

2. Define the map $\rho : M(p) \times M(p) \mapsto [0,\infty)$ as

$$\rho((D',U'),(D,U)) = \min_{G \in \mathcal{G}(p)} d_{SR}((GD'G^T, U'G^T),(D,U)).$$

Note that $\rho((D',U'),(D,U)) = \rho(\eta(G,(D',U')),(D,U))$ for any $G \in \mathcal{G}(p)$, meaning ρ is constant over orbits of its first variable, and $d_{\mathcal{PMSR}}(X,(D,U)) = \rho((D_X,U_X),(D,U))$, where $(D_X,U_X) \in \mathcal{F}^{-1}(X)$, for any $(X,(D,U)) \in S_p^{top} \times M(p)$. If we restrict ρ to the domain $M(p)^{top} \times M(p)$, where $M(p)^{top}$ denotes the eigen-decompositions of all members of S_p^{top}, this restriction of ρ induces a function on $S_p^{top} \times M(p)$ which is $d_{\mathcal{PMSR}}$.

The distance function d_{SR} is continuous on $M(p) \times M(p)$ since it is a metric on $M(p)$. Then ρ is also continuous on $M(p) \times M(p)$ because it is the minimum of a finite number of functions which are continuous on $M(p) \times M(p)$, which implies that the restriction of ρ to $M(p)^{top} \times M(p)$ is continuous on $M(p)^{top} \times M(p)$. Finally, since this restriction of ρ is continuous and is constant over orbits of its first variable, we have that its induced function $d_{\mathcal{PMSR}}$ is also continuous.

3. Pick any $X \in S_p^{top}$, $(D_0,U_0) \in M(p)$, and let $(D_n,U_n) \in M(p)$ be a sequence satisfying $d_{SR}((D_n,U_n),(D_0,U_0)) \to \infty$.

For any $n \geq 1$, one can choose $(D_X^{(n)}, U_X^{(n)}) \in \mathcal{F}^{-1}(X)$ that has minimal scaling-rotation distance from (D_n,U_n). By the Triangle Inequality, we have that

$$\begin{aligned}
&d_{SR}((D_n,U_n),(D_0,U_0)) \\
&\leq d_{SR}((D_n,U_n),(D_X^{(n)},U_X^{(n)})) + d_{SR}((D_X^{(n)},U_X^{(n)}),(D_0,U_0)) \\
&= d_{\mathcal{PMSR}}(X,(D_n,U_n)) + d_{SR}((D_X^{(n)}),(U_X^{(n)}),(D_0,U_0)) \\
&\leq d_{\mathcal{PMSR}}(X,(D_n,U_n)) + \max_{(D,U)\in\mathcal{F}^{-1}(X)} d_{SR}((D,U),(D_0,U_0))
\end{aligned}$$

Then we have $d_{\mathcal{PMSR}}(X,(D_n,U_n)) > d_{SR}((D_n,U_n),(D_0,U_0)) - B - 1$ for any $n \geq 1$, where $B = \max_{(D,U)\in\mathcal{F}^{-1}(X)} d_{SR}((D,U),(D_0,U_0))$ is a fixed positive constant which does not depend on n. Finally, we have that $d_{SR}((D_n,U_n),(D_0,U_0)) \to \infty$ implies that $d_{\mathcal{PMSR}}(X,(D_n,U_n)) \to \infty$.

□

Theorem 3.2.2. *Let $X, X_1, \ldots, X_n$ be i.i.d. random variables from a distribution on S_p^{top}. If the following conditions hold:*

1. *X has compact support or $\mathbf{E}[d^2_{\mathcal{PMSR}}(X,(D,U))] < \infty$ for any $(D,U) \in M(p)$.*
2. *$E^{(\mathcal{PMSR})} \neq \emptyset$.*
3. *$\cup_{n=1}^{\infty} E_n^{(\mathcal{PMSR})}$ enjoys the Heine-Borel property.*

then for every $\epsilon > 0$, there is a number $n = n(\varepsilon) > 0$ such that

$$\cup_{k=n}^{\infty} E_k^{(\mathcal{PMSR})} \subset \{(D,U) \in M(p) : d_{SR}(E^{(\mathcal{PMSR})},(D,U)) \leq \varepsilon\} \tag{3.5}$$

almost surely.

Proof. Conditions 1-3 are special cases of conditions from Theorems A.3 and A.4 from [27]. In [27], the coercivity condition also includes the requirement that there is a $(D_0,U_0) \in M(p)$ and $C > 0$ such that $Pr(d_{\mathcal{PMSR}}(X,(D_0,U_0)) < C) > 0$. We will first show that this property is satisfied as a consequence of condition 1 from Theorem 3.2.2.

From condition 1, it follows that $E[d^2_{\mathcal{PMSR}}(X,(D,U))] < \infty$ for any $(D,U) \in M(p)$. Since the support of X is a subset of S_p^{top}, we have $E[d^2_{\mathcal{PMSR}}(X,(I_p,I_p))] > 0$. Let $C = \sqrt{2}\sqrt{E[d^2_{\mathcal{PMSR}}(X,(I_p,I_p))]}$. Since $d_{\mathcal{PMSR}}(X,(I_p,I_p))$ is a non-negative real-valued random variable, it follows from the Markov Inequality that

$$P(d_{\mathcal{PMSR}}(X,(I_p,I_p)) < C) \geq 1 - \frac{E[d^2_{\mathcal{PMSR}}(X,(I_p,I_p))]}{C^2} = \frac{1}{2} > 0.$$

Theorems A.3 and A.4 also requires that $d_{\mathcal{PMSR}}$ be a continuous function on $S_p^{top} \times M(p)$, be uniformly continuous in its second argument, and possess the coercivity property in its second argument. These requirements for $d_{\mathcal{PMSR}}$ are not listed in the theorem conditions since they were verified in Lemma 3.2.2. □

Partial minimal scaling-rotation mean sets can be mapped to $\text{Sym}^+(p)$ in the following way:

Definition 3.2.2. The SPD sample and population partial scaling-rotation mean set are defined as

$$\mathcal{F}(E_n^{\mathcal{PMSR}}) = \{\mathcal{F}(D,U) : (D,U) \in E_n^{\mathcal{PMSR}}\}$$
$$\mathcal{F}(E^{\mathcal{PMSR}}) = \{\mathcal{F}(D,U) : (D,U) \in E^{\mathcal{PMSR}}\}.$$

The sets $\mathcal{F}(E_n^{\mathcal{PMSR}})$ and $\mathcal{F}(E^{\mathcal{PMSR}})$ are simply the images of the sample and population partial scaling-rotation mean sets under eigen-composition. Using these SPD versions of the sample and population partial minimal scaling-rotation mean sets, the version of strong consistency from Theorem 3.2.2 can be restated as convergence on $\text{Sym}^+(p)$:

Corollary 3.2.1. *Under the conditions of Theorem 3.2.2, it follows that for every $\varepsilon > 0$, there is a number $n = n(\varepsilon)$ such that*

$$\cup_{k=n}^{\infty}\mathcal{F}(E_k^{\mathcal{PMSR}}) \subset \{S \in Sym^+(p) : d_{MSR}(\mathcal{F}(E^{\mathcal{PMSR}}), S) \leq \varepsilon\} \tag{3.6}$$

with probability 1.

Proof. Let $A_{n,\varepsilon}$ denote the event from (3.5) and $B_{n,\varepsilon}$ denote the event from (3.6). We will show that if event $A_{n,\varepsilon}$ occurs, then event $B_{n,\varepsilon}$ must also occur. Fix $\varepsilon > 0$ and choose $n(\varepsilon)$ so that for $n \geq n(\varepsilon)$, event $A_{n,\varepsilon}$ occurs with probability 1.

Pick $n = n(\varepsilon)$ and choose $S \in \cup_{k=n}^{\infty}\mathcal{F}(E_k^{\mathcal{PMSR}})$. From the definition of the SPD sample partial minimal scaling-rotation mean set, there is some $(D,U) \in \cup_{k=n}^{\infty}E_k^{\mathcal{PMSR}}$ such that $S = \mathcal{F}(D,U)$. Likewise, if $S^* \in \mathcal{F}(E^{(\mathcal{PMSR})})$, there is some $(D^*,U^*) \in E^{(\mathcal{PMSR})}$ such that $S^* = \mathcal{F}(D^*,U^*)$. By the definition of the minimal scaling-rotation distance, we have that

$$d_{MSR}(S,S^*) \leq d_{SR}((D,U),(D^*,U^*))$$

for any $S^* \in \mathcal{F}(E^{(\mathcal{PMSR})})$, which implies that

$$d_{MSR}(S,\mathcal{F}(E^{(\mathcal{PMSR})})) = \inf_{S^* \in \mathcal{F}(E^{(\mathcal{PMSR})})} d_{MSR}(S,S^*)$$
$$\leq \inf_{(D^*,U^*) \in E^{(\mathcal{PMSR})}} d_{SR}((D,U),(D^*,U^*)) = d_{SR}((D,U),E^{(\mathcal{PMSR})}) \leq \varepsilon.$$

Thus, we have that if event $A_{n,\varepsilon}$ occurs, then event $B_{n,\varepsilon}$ will occur, which implies that $P(B_{n,\varepsilon}) \geq P(A_{n,\varepsilon}) = 1$. □

A usual way to define a Central Limit Theorem for an estimator that lives on a Riemannian manifold of dimension d is to first define a chart, a locally smooth and invertible function which maps the manifold to $\mathbb{R}^d$, and then establish the asymptotic distribution of the linearized estimator. We use the following definition of a Central Limit Theorem for a Riemannian manifold valued estimator from [26]:

Definition 3.2.3. Let P be a D-dimensional Riemannian manifold. A P-valued estimator μ_n of population mean $\mu \in P$ satisfies a CLT if in any local chart (ϕ, U) near $\mu = \phi^{-1}(0)$ there are a suitable $D \times D$ matrix A_ϕ and a Gaussian random variable G_ϕ with zero mean and symmetric semi-definite covariance matrix Σ_ϕ such that

$$\sqrt{n} A_\phi(\phi(\mu_n) - \phi(\mu)) \to G_\phi$$

in distribution as $n \to \infty$.

Our parameter space of interest $M(p)$ is a Riemannian manifold of dimension $p + \frac{(p-1)p}{2}$. We choose the following local chart ϕ of $M(p)$ centered at $\phi^{-1}(0) = (D, U)$:

$$\phi_{(D,U)}((\Lambda, R)) = (Log(\Lambda D^{-1}), Log(RU^T)), \Lambda \in Diag^+(p), R \in SO(p) \tag{3.7}$$

$$\phi^{-1}_{(D,U)}((L, A)) = (Exp(L)D, Exp(A)U), L \in Diag(p), A \in \mathfrak{so}(p) \tag{3.8}$$

Theorem 3.2.3. *Suppose that (D, U) is a member of $E^{(\mathcal{PMSR})}$ which is unique up to the action of $\mathcal{G}(p)$ on $M(p)$. Assume that $(D_n, U_n) \in E_n^{(\mathcal{PMSR})}$ is unique up to the action of $\mathcal{G}(p)$ and is a strongly consistent estimator of (D, U) in the sense of Theorem 3.2.2. If X has compact support such that*

$$d_{\mathcal{PMSR}}(X, (D, U)) < \frac{\sqrt{K}}{4} \beta_{\mathcal{G}(p)}$$

with probability 1, then for any choice $(D'_n, U'_n) \in E_n^{(\mathcal{PMSR})}$, there is a sequence $G_n \in \mathcal{G}(p)$ such that $(D_n, U_n) = \eta(G_n, (D'_n, U'_n))$ satisfies a CLT using the chart defined in (3.7).

Proof. This is a special case of Theorem 6 from [26]. That theorem has a condition that the distance function be continuous on its product space, which we do not include as a condition for Theorem 3.2.3 since the continuity of $d_{\mathcal{PSR}}$ on $S_p^{top} \times M(p)$ is demonstrated in Lemma 3.2.2.

Theorem 6 from [26] also has a condition which states, in the context of Theorem 3.2.3, that for any choice $(D_n', U_n') \in E_n^{\mathcal{PMSR}}$, there is a sequence $G_n \in \mathcal{G}(p)$ such that $(D_n, U_n) = \eta(G_n, (D_n', U_n'))$ is a consistent estimator of (D, U). We do not include this as a condition for Theorem 3.2.3 since this result follows from the assumption that $(D_n, U_n) \in E_n^{(\mathcal{PMSR})}$ is unique up to the action of $\mathcal{G}(p)$ and is a strongly consistent estimator of (D, U). Note that since (D_n, U_n) is unique up to the action of $\mathcal{G}(p)$, any selection $(D_n', U_n') \in E_n^{\mathcal{PMSR}}$ belongs to the same orbit as (D_n, U_n), meaning there exists a $G_n \in \mathcal{G}(p)$ such that $(D_n, U_n) = \eta(G_n, (D_n', U_n'))$.

The only remaining condition from Theorem 6 of [26] that we need to verify is the smoothness of the map

$$(L, A) \mapsto d^2_{\mathcal{PMSR}}(X, \phi^{-1}_{(D,U)}((L, A))) \tag{3.9}$$

where $A \in \mathfrak{so}(p)$ and $L \in Diag(p)$. We need to show that there is some neighborhood about $(0, 0)$ such that the map is twice continuously differentiable with probability 1 when (L, A) belongs to that neighborhood.

There are two possible sources of non-differentiability in (3.9): (i) changing of the eigen-decomposition of X that forms a minimal pair with $\phi^{-1}_{(D,U)}((L, A))$ as (L, A) varies within some neighborhood about $(0, 0)$, and (ii) the fact that $d_{SO(p)}(R', R)$ is not differentiable with respect to R when R belongs to the cut locus of R'.

We first restrict (L, A) to belong to a neighborhood centered at $(0, 0)$ so that for any X in the support of the distribution, the eigen-decomposition of X that forms a minimal pair with $\phi^{-1}_{(D,U)}((L, A))$ does not change as (L, A) varies within that neighborhood. Based on the assumption that $d_{\mathcal{PMSR}}(X, (D, U)) < \sqrt{K}\beta_{\mathcal{G}(p)}/4$ with probability 1, we will restrict (L, A) to belong to a neighborhood centered at $(0, 0)$ so that $d_{SR}(\phi^{-1}_{(D,U)}((L, A)), (D, U)) < \sqrt{K}\beta_{\mathcal{G}(p)}/4$. Suppose that $(D_X, U_X) \in \mathcal{F}^{-1}(X)$ forms a minimal pair with (D, U). By the Triangle Inequality, we have $d_{SR}(\phi^{-1}_{(D,U)}((L, A)), (D_X, U_X)) < \sqrt{K}\beta_{\mathcal{G}(p)}/2$ for any (L, A) belonging to

that neighborhood. By Lemma 3.2.1, no other eigen-decomposition of X can belong to that neighborhood centered at $\phi^{-1}_{(D,U)}((L,A))$, implying that (D_X, U_X) and $\phi^{-1}_{(D,U)}((L,A))$ form a minimal pair. Then we have $d^2_{\mathcal{PMSR}}(X, \phi^{-1}_{(D,U)}((L,A))) = d^2_{SR}((D_X,U_X), \phi^{-1}_{(D,U)}((L,A)))$ for any (L,A) such that $d_{SR}(\phi^{-1}_{(D,U)}((L,A)),(D,U)) < \sqrt{K}\beta_{\mathcal{G}(p)}/4$ with probability 1.

From the definitions of the partial minimal scaling-rotation distance, scaling-rotation distance and (3.8), we have that

$$\begin{aligned} & d^2_{\mathcal{PMSR}}(X, \phi^{-1}_{(D,U)}((L,A))) \\ & = d^2_{SR}((D_X,U_X), \phi^{-1}_{(D,U)}((L,A))) \\ & = d^2_{\mathcal{D}^+(p)}(D_X, Exp(L)D) + K d^2_{SO(p)}(U_X, Exp(A)U) \end{aligned} \tag{3.10}$$

for any (L,A) such that $d_M(\phi^{-1}_{(D,U)}((L,A)),(D,U)) < \sqrt{K}\beta_{\mathcal{G}(p)}/4$ with probability 1. We next address restrictions on (L,A) that will guarantee that (3.10) will be twice continuously differentiable.

The map $L \mapsto d^2_{\mathcal{D}^+(p)}(D_X, Exp(L)D)$ is continuously differentiable of all orders at any $L \in Diag(p)$ since it can be shown that $d^2_{\mathcal{D}^+(p)}(D_X, Exp(L)D)$ is a quadratic function in each diagonal coordinate of L.

The map $A \mapsto d^2_{SO(p)}(U_X, Exp(A)U)$ will not be twice continuously differentiable for all $A \in \mathfrak{so}(p)$. More specifically, it will fail to be differentiable if and only if there is not a unique geodesic on $SO(p)$ connecting U_X and $Exp(A)U$. Since $d^2_{SO(p)}(U_X, Exp(A)U) = d^2_{SO(p)}(I_p, Exp(A)UU_X^T) = \frac{1}{2}\|Log(Exp(A)UU_X^T)\|^2_F$, this question can be reduced to assessing the uniqueness of $Log(Exp(A)UU_X^T)$. A unique principal logarithm of $Exp(A)UU_X^T$ exists if none of the eigenvalues of $Exp(A)UU_X^T$ lie on the negative real line ([25]).

We next explain an equivalent sufficient condition for the existence of a unique principal matrix logarithm of a rotation matrix. Let $k = \lfloor \frac{p}{2} \rfloor$, the floor of $\frac{p}{2}$. Recall that for any $R \in SO(p)$, its eigenvalues lie on the unit circle of the complex plane. If p is even, its eigenvalues split off into k complex conjugate pairs: $e^{i\theta_1}, e^{-i\theta_1}, \ldots, e^{i\theta_k}, e^{-i\theta_k}$. If p is odd, R has an eigenvalue equal to 1 and its remaining $p-1$ eigenvalues form k complex conjugate pairs $e^{i\theta_1}, e^{-i\theta_1}, \ldots, e^{i\theta_k}, e^{-i\theta_k}$. Then R will have an eigenvalue on the negative real line if and only if an angle from one of its complex conjugate pairs is equivalent to π modulo 2π. Thus, R will have a unique principal matrix logarithm if $\theta_j \not\equiv \pi \pmod{2\pi}$ for $j = 1, \ldots, k$.

For $R \in SO(p)$ with complex conjugate eigenvalues $e^{i\theta_1}, e^{-i\theta_1}, \ldots, e^{i\theta_k}, e^{-i\theta_k}$, it can be shown that $d_{SO(p)}(R, I_p) = (\sum_{j=1}^{k} \tilde{\theta}_j^2)^{1/2}$ where $\tilde{\theta}_j \equiv \theta_j \pmod{2\pi}$ and $|\tilde{\theta}_j| \leq \pi$ for $j = 1, \ldots, k$. Then we have that $d_{SO(p)}(R, I_p) < \pi$ implies that $\theta_j \not\equiv \pi \pmod{2\pi}$ for $j = 1, \ldots, k$, which implies that R has a unique principal matrix logarithm.

Since $\beta_{\mathcal{G}(p)} \leq \pi/2$ for any $p \geq 2$, assuming that $d_{\mathcal{PMSR}}(X, (D, U); K) < \sqrt{K}\beta_{\mathcal{G}(p)}/4$ with probability 1 and (L, A) satisfies $d_{SR}(\phi^{-1}_{(D,U)}((L, A)), (D, U); K) < \sqrt{K}\beta_{\mathcal{G}(p)}/4$ implies $d_{SR}((D_X, U_X), \phi^{-1}_{(D,U)}((L, A))) < \sqrt{K}\pi$ with probability 1, from which it follows that $d_{SO(p)}(U_X, Exp(A)U) < \pi$ and $Exp(A)UU_X^T$ has a unique principal logarithm with probability 1.

Thus, we have that restricting (L, A) to a neighborhood centered at $(0, 0)$ such that $d_M(\phi^{-1}_{(D,U)}((L, A)), (D, U)) < \sqrt{K}\beta_{\mathcal{G}(p)}/4$ implies that (3.9) will be twice continuously differentiable with probability 1. □

3.3 DTI ANALYSIS INTERPOLATION EXAMPLES

In this section we illustrate the usefulness of scaling-rotation mean estimation for interpolation of diffusion tensors along one and two dimensions. We compare interpolation via the scaling-rotation framework to interpolation using other widely-used geometric frameworks for $Sym^+(p)$, including the Log-Euclidean, Affine-Invariant, and Procrustes Size-and-Shape frameworks.

Diffusion tensor imaging provides measurements of local water diffusion along white matter fiber tracts at voxels within a biological object [46]. The diffusion tensor model assumes that this movement of water molecules at time t follows a mean-zero trivariate normal distribution with covariance matrix $2\Sigma t \in Sym^+(3)$ [1]. The quantity Σ is the diffusion tensor.

Two core concepts for describing diffusion tensors are **anisotropy** and **isotropy**. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$ denote the eigenvalues of a diffusion tensor. The case $\lambda_1 = \lambda_2 = \lambda_3$ describes a perfectly isotropic diffusion tensor, while the case $\lambda_1 \gg \lambda_2, \lambda_3$ describes an anisotropic diffusion tensor. A voxel of the brain that is modeled by an isotropic diffusion

tensor can be thought of as a region where water molecules move with no preferred direction, while a voxel with a highly anisotropic diffusion tensor can be thought of as a region where the movement of water molecules occurs primarily along one direction.

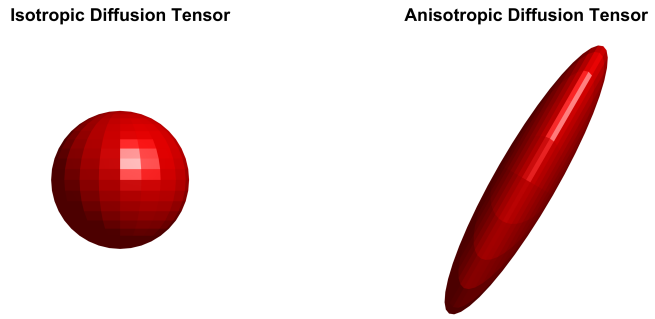

Figure 7: Isotropic (left) and highly anisotropic (right) diffusion tensors.

The eigen-decomposition of a diffusion tensor plays an important role in DTI analysis. Perhaps the simplest descriptive measure of the strength of diffusivity at a voxel is the Mean Diffusivity (MD):

$$MD(D) = \frac{1}{3}tr(D) = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}. \tag{3.11}$$

One of the most commonly used measures of anisotropy of a diffusion tensor D is Fractional Anisotropy (FA) [31]

$$FA(D) = \sqrt{\frac{3[(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2]}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}, \tag{3.12}$$

where $\bar{\lambda} = MD(D)$. $FA(D)$ ranges from 0 in the case of perfect isotropy to 1 in the case of linear anisotropy ($\lambda_1 > \lambda_2 = \lambda_3 = 0$). A closely related measure of anisotropy is Procrustes Anisotropy (PA), which computes the fractional anisotropy of $\sqrt{\lambda_1}, \sqrt{\lambda_2}$, and $\sqrt{\lambda_3}$ (i.e. $PA(D) = FA(\sqrt{D})$) [15].

3.3.1 Common Frameworks for Sym-Plus(p)

In this subsection we review some established geometric frameworks for $Sym^+(p)$, including a Eucidean framework, a Log-Euclidean framework ([2]), a Riemannian affine-invariant framework (citations needed), and matrix square root based approaches including a Cholesky-based framework, a root-Euclidean framework, and a framework based on the Procrustes Size-and-Shape metric of [15].

We first discuss a Euclidean framework for $Sym^+(p)$. Let $X, Y \in Sym^+(p)$. Since $Sym^+(p) \subset Sym(p)$, which is isomorphic to $\mathbb{R}^{\frac{p(p+1)}{2}}$ and can be metrized with the Frobenius norm $\|.\|_F$, one can can measure the distance between X and Y via the distance function

$$d_E(X, Y) = \|X - Y\|_F.$$

The distance $d_E(X, Y)$ measures the length of the line

$$\chi_E(t) = (1 - t)X + tY$$

for $t \in [0, 1]$.

We next move on to the Log-Euclidean framework for $Sym^+(p)$. Again, let $X, Y \in Sym^+(p)$. Recall that $Exp : Sym(p) \mapsto Sym^+(p)$ is bijective, so any $M \in Sym^+(p)$ can be uniquely identified with $Log(M) \in Sym(p)$. Hence, $Sym^+(p)$ can be thought of as a "log-linear" space, and one can measure the distance between $X, Y \in Sym^+(p)$ via the distance function

$$d_{LE}(X, Y) = \|Log(X) - Log(Y)\|_F.$$

Geodesics corresponding to $d_{LE}(.,.)$ are curves of the form

$$\chi_{LE}(t) = Exp((1 - t)Log(X) + tLog(Y))$$

for $t \in [0, 1]$. We refer the reader to [2] for a more detailed account of the geometric properties of the Log-Euclidean characterization of $Sym^+(p)$.

We finally present an affine-invariant Riemannian framework for $Sym^+(p)$. Let $X, Y \in Sym^+(p)$. Note that we can write Y as $Y = M_X(M_X^{-1}Y(M_X^{-1})^T)M_X^T$, where $M_X \in GL(p)$

is a square root of X (i.e. $M_X M_X^T = X$) and $M_X^{-1} Y (M_X^{-1})^T \in Sym^+(p)$ since $Sym^+(p)$ is closed under conjugation by matrices in $GL(p)$. For $t \in [0,1]$, the curve

$$\chi_{AI}(t) = M_X Exp(Log(M_X^{-1} Y (M_X^{-1})^T) t) M_X^T$$

traces a geodesic connecting X and Y under the usual Riemannian inner product for $Sym^+(p)$ that is invariant under the group action. This curve has length

$$d_{AI}(X,Y) = \|Log(M_X^{-1} Y (M_X^{-1})^T)\|_F.$$

It is important to note that the curve $\chi_{AI}(t)$ and its corresponding distance function $d_{AI}(.)$ are invariant with respect to the choice of square root for X. For more information on this affine-invariant Riemannian framework, we refer the reader to [37].

We next present frameworks based on square root decompositions, beginning with the Cholesky decomposition. Recall that the Cholesky decomposition uniquely factors a matrix $X \in Sym^+(p)$ as $X = QQ^T$, where Q is a lower triangular matrix with positive diagonal entries. Given $X, Y \in Sym^+(p)$, suppose that X and Y have Cholesky decompositions $X = Q_X Q_X^T$ and $Y = Q_Y Q_Y^T$. Via Cholesky composition, one can define a smooth curve on $Sym^+(p)$ running from X to Y using the formula

$$\chi_C(t) = [Q_X + t(Q_Y - Q_X)][Q_X + t(Q_Y - Q_X)]^T$$

for $t \in [0,1]$. This curve traces a geodesic on $Sym^+(p)$ from X to Y with length

$$d_C(X,Y) = \|Q_Y - Q_X\|_F.$$

Another unique square root decomposition of an SPD matrix is the symmetric square root decomposition. Given $X \in Sym^+(p)$ with eigen-decomposition $X = UDU^T$, we define its symmetric square root as $X^{1/2} = UD^{1/2}U^T$, where $D^{1/2}$ is a diagonal matrix whose diagonal entries are the positive square roots of the diagonal entries of D. Given $X, Y \in Sym^+(p)$, one can create the following curve

$$\chi_H(t) = [X^{1/2} + t(Y^{1/2} - X^{1/2})]^2$$

which traces a geodesic on $Sym^+(p)$ from X to Y for $t \in [0,1]$ with length

$$d_H(t) = \|Y^{1/2} - X^{1/2}\|_F.$$

This geometric framework for $Sym^+(p)$ based on the symmetric square root is referred to as the Root-Euclidean framework in [48].

Note that the Cholesky decomposition and symmetric square root decomposition methods yield two out of infinitely many possible square roots of an SPD matrix. Indeed, if $X \in Sym^+(p)$ and $X = LL^T$ then LR, where $R \in O(p)$, is also a square root of X since $(LR)(LR)^T = X$. Given $X, Y \in Sym^+(p)$ with square roots Q_X and Q_Y, one can define a measure of distance between X and Y as

$$d_S(X,Y) = \inf_{R \in O(p)} \|Q_X - Q_Y R\|_F,$$

which finds the square root of Y that is closest to the initial square root Q_X of X under the Frobenius norm. Given initial square roots Q_X and Q_Y for X and Y, respectively, it can be shown that

$$\hat{R} = \arg \inf_{R \in O(p)} \|Q_X - Q_Y R\|_F = UW^T$$

where $W, U \in O(p)$ come from the singular value decomposition $Q_X Q_Y^T = W \Lambda U^T$. A geodesic from X to Y on $Sym^+(p)$ with respect to this metric will be of the form

$$\chi_S(t) = [Q_X + t(Q_Y \hat{R} - Q_X)][Q_X + t(Q_Y \hat{R} - Q_X)]^T$$

for $t \in [0,1]$. More information on this framework can be found in [15].

Like the Scaling-Rotation framework, the Log-Euclidean and affine-invariant Riemannian frameworks for $Sym^+(p)$ are also intrinsic frameworks since the geodesics $\chi_{LE}(t)$ and $\chi_{AI}(t)$ lie in $Sym^+(p)$ for all $t \in \mathbb{R}$. While the curve $\chi_E(t)$ lies in $Sym(p)$ for all $t \in \mathbb{R}$, there will be some t for which $\chi_E(t)$ will fail to be positive-definite. The curves $\chi_C(t)$, $\chi_H(t)$, and $\chi_S(t)$ will be non-negative definite, symmetric matrices for all $t \in \mathbb{R}$, but can be less than full rank for some t. Hence, the Euclidean, Cholesky, Root-Euclidean, and Procrustes frameworks are all extrinsic frameworks for $Sym^+(p)$.

Of all the frameworks discussed in this section, the Procrustes size-and-shape framework perhaps shares the most conceptual similarities with the Scaling-Rotation framework. Both

frameworks parameterize $Sym^+(p)$ via non-unique matrix decompositions (matrix square root and eigen-decomposition), and the distance between elements $X, Y \in Sym^+(p)$ is defined as the minimum distance between all possible decompositions of X and Y.

3.3.2 Interpolation of Diffusion Tensors

The interpolation of diffusion tensors is vital to tracking white matter fiber tracts in the brain and reducing noise in image data ([4],[10]). In this section we compare interpolation between two diffusion tensors under the minimal scaling-rotation framework and the geometric frameworks described in the previous section. We analyze interpolation under three deformation patterns: scaling plus rotation, pure rotation, and pure scaling from an anisotropic diffusion tensor to an isotropic diffusion tensor. For each of the examples, we have plots of 11 equally spaced points $\left(t = \frac{j}{10} \text{ for } j = 0, 1, \ldots, 10\right)$ along a geodesic curve and a comparison of MD, FA, PA, log determinant, and rotation angle (defined as the angle of swing from the major axis at time 0 to the major axis at time t) among the geometric frameworks.

3.3.3 Example I: Interpolation Under Scaling Plus Rotation

For our first example, we analyze interpolation from X_1 to Y_1, where

$$X_1 = \begin{pmatrix} 15 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } Y_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 100 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

As can be seen in Figure 8, the "pointy" shape of X_1 and Y_1 is preserved along the geodesic curve under the Scaling-Rotation, Cholesky, and Procrustes frameworks. This is reflected in the plots of FA and PA, in which we see that only those three interpolation types exhibit monotone growth of FA and PA as X_1 deforms to Y_1. The Root Eucliean and Euclidean interpolants appear to suffer most from the "fattening effect", which describes interpolated diffusion tensors that are more isotropic than the endpoints of a geodesic curve [10]. Only the Log-Eucliean interpolation path exhibits the "shrinking effect" [4] (when some interpolated diffusion tensors have lower MD than the endpoints of a geodesic curve) for this example.

The Log-Euclidean and Scaling-Rotation frameworks both possess log-linear growth of determinant since $\chi_{LE}(t)$ and χ_{MSR} both satisfy

$$\log(\det(\chi_{LE}(t))) = \log(\det(\chi_{MSR}(t))) = (1-t)\log(\det(X_1)) + t\log(\det(Y_1)).$$

In addition to the "fattening effect," the Euclidean and Root Euclidean frameworks both also exhibit the "swelling effect", which occurs when some interpolated diffusion tensors have larger determinant than the endpoints of the geodesic curve [2].

Only the Scaling-Rotation interpolation exhibits linear evolution of rotation angle along its geodesic path. In fact, the rotation angle at any point along the Scaling-Rotation geodesic path from X to Y is given by

$$\theta_{MSR}(t) = \frac{\|Log(U_Y U_X^T)\|_F}{\sqrt{2}} t$$

where (D_X, U_X) and (D_Y, U_Y) are eigen-decompositions of X and Y. The plot of rotation angles in Figure 9 shows that the other interpolated diffusion tensors rotate from the orientation of X_1 to the orientation Y_1 at variable speeds, with all of the frameworks except the Log-Euclidean framework rotating with faster initial speed than the Scaling-Rotation framework.

All three of the Scaling-Rotation, Cholesky, and Procrustes interpolation paths exhibit monotone growth of FA and PA from X_1 to the more anisotropic Y_1, and none of them suffer from the swelling, shrinking, or fattening effects. Interpolation via the Cholesky or Procrustes frameworks yields visually similar geodesic paths from X_1 to Y_1 that possess similar DTI summary measures, as seen in Figure 8 and Figure 9, respectively. It could be argued that the Scaling-Rotation method of interpolation is desirable over the Cholesky and Procrustes methods since it is the only framework in which interpolants rotate linearly, which is consistent with human visual perception of rotation.

3.3.4 Example II: Interpolation Under Pure Rotation

For our second example, we analyze deformation from X_2 to Y_2, where

$$X_2 = \begin{pmatrix} 15 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } Y_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 15 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Visualization of the interpolation curves and DTI summary statistics for this example can be found in Figure 10 and Figure 11. Scaling-Rotation interpolation was the only method that perfectly preserved the shape of X_2 as it rotated to align with Y_2, as confirmed by its constant FA and PA curves. Diffusion tensors along the Euclidean and Affine-Invariant interpolation paths appear to suffer most from the "fattening effect," especially near the midpoints ($t = 0.5$) of the geodesic curves. All interpolation methods except for the Scaling-Rotation and Euclidean methods suffer from the "shrinking effect" in this example. The Scaling-Rotation interpolation method was also one of the few interpolation types (including the Log-Euclidean and Affine-invariant methods) to maintain a constant determinant along its geodesic path from X_2 to Y_2. Finally, Scaling-Rotation interpolation is the only method possessing linear evolution of rotation angle, while the other frameworks provide nearly linear patterns of rotation during interpolation.

In summary, the Scaling-Rotation interpolation method outperforms the other six interpolation types in the comparisons of DTI summary values in Figure 11, and is the only interpolation method to preserve volume and shape, or in other words acting as pure rotation, along its geodesic path in this example. The Cholesky and Procrustes interpolation methods yield similar geodesic paths to the Scaling-Rotation method, and seem to yield interpolants that are close in volume and degree of anisotropy to endpoints X_2 and Y_2, making them good alternative options to the Scaling-Rotation method for this example.

3.3.5 Example III: Interpolation Under Pure Scaling

For our third example, we analyze interpolation from X_3 to isotropic Y_3, where

$$X_3 = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 16 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \text{ and } Y_3 = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

Visualization of the interpolation curves and plots of DTI summary measures can be found in Figure 12 and Figure 13. Since deformation from X_3 to Y_3 consists of pure scaling, the interpolation curves for the Scaling-Rotation, Log-Euclidean, and Affine-Invariant methods are equivalent [28]. All of the interpolation methods exhibit monotone decreasing FA and PA curves. The Cholesky, Procrustes, and Euclidean interpolation types all suffer from the "swelling effect" for this example. For all interpolation methods except for the Cholesky method, the rotation angle equals 0 for all t (i.e. the orientation of the interpolants matches that of X_2), while the orientation of the Cholesky interpolants rotates by about 15 degrees along the interpolation curve.

3.3.6 Example IV: 2D Interpolation I

In this first 2D interpolation example, the weighted sample means of the four corner points of the 5×5 grid are computed to interpolation between the corner points and fill in the grid. The upper left corner point is an isotropic diffusion tensor and the three remaining corner points are highly anisotropic and differ only by rotations. All four corner points have the same determinant. The top left, top right, bottom left, and bottom right have coordinates $(1,1), (1,5), (5,1)$, and $(5,5)$, respectively, and the weights used to compute the interpolants at an interior point are the reciprocals of the squared distances from the interior coordinate to the corner coordinate points, with the weights scaled to sum to 1. The colors of the diffusion tensor range from yellow for FA of 0 to red for FA of 1.

Of all the methods shown in Figure 14, the scaliing-rotation framework does the best job of preserving shape and anisotropy along paths connecting corner points that differ only by rotations. The other frameworks exhbit a smoother progression of anisotropy along paths

emanating from the upper left corner compared to the scaling-rotation framework. Not surprisingly, the Euclidean interpolants appear to suffer from the swelling effect compared to the other frameworks.

3.3.7 Example V: 2D Interpolation II

Interpolation in this 2D example, which can be seen in Figure 15, were performed as in Example IV. The top left corner point is an isotropic diffusion tensor, the top-right corner point is prolate diffusion tensor (it has a maximum eigenvalue and the two smaller eigenvalues are equal), and the bottom corner points are highly anisotropic and only differ by rotation. The Scaling-Rotation, Log-Euclidean, and Procrustes Size-and-Shape interpolants look similar, with the Procrustes interpolants generally having larger determinants compared to those from Log-Euclidean and Scaling-Rotation interpolation. As in Example IV, Euclidean 2D interpolation yields interpolants that generally have larger determinants compared to the interpolants using the other three frameworks.

3.3.8 Summary of Examples I-V

The Scaling-Rotation interpolation method performed well in comparison to the other methods described in this section. In all three examples, the Scaling-Rotation interpolation method never suffered from the swelling, fattening, or shrinking effects and always yielded monotone evolution of FA and PA. In examples I and II, the Cholesky and Procrustes frameworks provide good alternatives to the Scaling-Rotation method. In example III, the Scaling-Rotation, Log-Euclidean, and Affine-Invariant methods yields equivalent interpolation curves.

It should be noted that Cholesky framework is not used in practice for analyzing diffusion tensors since the Cholesky metric is not invariant to change of coordinates. More specifically, for $X_1, X_2 \in Sym^+(p)$, it is possible to have

$$d_C(X_1, X_2) \neq d_C(RX_1R^T, RX_2R^T)$$

for some $R \in SO(p)$. An illustration of this problem can be found in section 4.3.3 of [48].

In the 2D interpolation examples, we saw that the Scaling-Rotation framework performed well at yielding nearly pure rotation paths between corner points which differ only by rotation, while the Log-Euclidean and Procrustes Size-and-Shape frameworks were better at yielding nearly pure scaling paths between corner points which differ only by scaling.

3.4 DATA APPLICATIONS

3.4.1 Multivariate Tensor-Based Morphometry

In [38], the authors compared the lateral ventricular structure in the brains of 17 preterm and 19 fullterm infant children. After an MRI scan of a subject's brain was obtained and processed through an image processing pipeline, the shape data collected at 102816 vertices on the surfaces of their left and right ventricles were mapped onto the left and right ventricles of a template brain image, after which the 2×2 Jacobian matrix J from that surface registration transformation was computed at each vertex for each subject. The deformation tensor $X = (JJ^T)^{1/2} \in Sym^+(2)$ was then computed at each vertex for each subject. To summarize the structure of the data, there are 102816 vertices along the surfaces of the template ventricles, and at each vertex there are deformation tensors (2×2 SPD matrices) from 17 preterm and 19 full term infants.

One way that the authors tested for differences in ventricular shape between the fullterm and preterm infants was by performing two sample location tests at each vertex via use of the Log-Euclidean version of Hotelling's T^2 test statistic introduced in [33]. Conceptually, this procedure "linearizes" the deformation tensors via the Log-Euclidean framework and then computes the classic T^2 test statistic with the linearized data.

Similarly, one could compute a scaling-rotation version of Hotelling's T^2 test statistic by linearizing the deformation tensors at a given vertex using the scaling-rotation framework and then computing the test statistic using the linearized data. To map an observation X to its scaling-rotation tangent space centered at (D, U) we will use the map $\tau_{(D,U)} : Sym^+(p) \mapsto$

$\times Diag(p) \times \mathfrak{so}(p)$ given by

$$\tau_{(D,U)}(X) = (Log(D_X^* D^{-1}), Log(U_X^* U^{-1})),$$

where (D_X^*, U_X^*) is an eigen-decomposition of X that has minimal geodesic distance from (D, U). We propose the following procedure for computing the scaling-rotation Hotelling's T^2 test statistic given observations $X_1, \ldots, X_{n_1}$ from sample 1 of size n_1 and observations $Y_1, \ldots, Y_{n_2}$ from sample 2 of size n_2:

1. Compute $(\hat{D}_1, \hat{U}_1)$ and $(\hat{D}_2, \hat{U}_2)$, sample partial scaling-rotation means of sample 1 and 2.
2. Let (D_1^*, U_1^*) denote the eigen-decomposition of $\hat{\Sigma}_1 = \hat{U}_1 \hat{D}_1 \hat{U}_1^T$ which has minimal scaling-rotation distance from $(\hat{D}_2, \hat{U}_2)$ and compute $c = vec(\tau_{(\hat{D}_2, \hat{U}_2)}(\hat{\Sigma}_1))$.
3. For $i = 1, \ldots, n_1$ compute $x_i = vec(\tau_{(D_1^*, U_1^*)}(X_i)) + c$ and for $i = 1, \ldots, n_2$ compute $y_j = vec(\tau_{(\hat{D}_2, \hat{U}_2)}(Y_j))$.
4. Compute Hotelling's T^2 test statistic using the linearized data $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$.

As an example, we have computed the Log-Euclidean and scaling-rotation versions of Hotelling's T^2 test statistic at vertex 75412 from the dataset used by [38]. The scaling-rotation T^2 statistic is 37.38 while the Log-Euclidean T^2 is 0.99. To investigate this discrepancy in test statistics, we have plotted the linearized versions of the deformation tensors using the Log-Euclidean and scaling-rotation frameworks in the top plots in Figure 16. Observations from preterm infants are blue and observations from fullterm infants are red. There is little separation between the fullterm and preterm observations in the Log-Euclidean tangent space, whereas there is almost perfect separation between the fullterm and preterm observations along the angle direction in the scaling-rotation tangent space. To help see the separation between the fullterm and preterm observations in the scaling-rotation tangent space more clearly, we have plotted the log-eigenvalues in the bottom left plot and the angles as points on the unit circle in the bottom right plot of Figure 16. These bottom plots show a lot of overlap in log-eigenvalues between fullterm and preterm infants and very little overlap of eigenvector rotation angles between fullterm and preterm infants, suggesting that the main location differences between the two groups at vertex 75412 are due to rotation.

With this example in mind, more work is needed to investigate if the Log-Euclidean T^2 statistic is less sensitive in detecting rotation differences between groups than the scaling-rotation T^2 statistic, and a full vertex-wise analysis of the ventricle data from [38] using the scaling-rotation T^2 statistic should be carried out to see if the scaling-rotation T^2 statistic detects new and scientifically meaningful regions of significant structural differences between the preterm and fullterm groups.

3.4.2 Diffusion Tensor Interpolation

We next present an application of the scaling-rotation estimation framework to interpolation of diffusion tensor imaging (DTI) data. DTI is a modality of magnetic resonance imaging that models local water diffusion along white matter fiber tracts within a biological object, and has been used extensively for brain imaging [46]. The diffusion tensor model assumes that the movement of water molecules at time t follows a mean-zero trivariate normal distribution with covariance matrix $2\Sigma t \in Sym^+(3)$, where Σ is referred to as the diffusion tensor [1].

Weighted averaging of diffusion tensors is useful for image interpolation, fiber tracking [4], and smoothing to removing noise inherent in the image generation process [9]. Using data from [47], we interpolate the diffusion tensors from an image of a coronal slice of a healthy human brain to improve image resolution. Interpolation of diffusion tensors in two dimensions can be done by subdividing each collection of four neighboring diffusion tensors into a grid of equally space points with the original diffusion tensors placed at the four corners. Weighted averaging of the original diffusion tensors can then be done to add interpolated diffusion tensors to the non-corner points on the grid. Weights are usually chosen as functions of the distances between the non-corner points and the locations of the nearest diffusion tensors from the original data.

In Figure 17 we have plotted Fractional Anisotropy (FA) maps using the original diffusion tensors and the interpolated diffusion tensors. The fractional anisotropy of a diffusion tensor D is computed as

$$FA(D) = \sqrt{\frac{3}{2}}\sqrt{\frac{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}},$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$ are the eigenvalues of D. $FA(D)$ describes the shape of D, ranging from 0 when $\lambda_1 = \lambda_2 = \lambda_3$ (D is isotropic) to 1 when $\lambda_1 > \lambda_2 = \lambda_3 = 0$ (D is perfectly anisotropic). Values of FA near 0 in a region suggeset minimal structure (i.e. water molecues diffuse easily in all directions), while regions where FA is close to 1 are highly structured (i.e. water molecules tend to diffuse along the direction of the white matter structures). The colors in the FA maps range from black for FA=0 to white for FA=1. When comparing the FA maps of the original diffusion tensors and the interpolated diffusion tensors, we can see that interpolation greatly improves the resolution of the FA map.

3.5 DISCUSSION

In this chapter we have presented the first statistical estimation methods for $Sym^+(p)$ based on the scaling-rotation framework of [28], which sets the foundation for the development of more scaling-rotation statistical methods, such as two sample hypothesis tests of location differences, hypothesis testing for a variety of eigenvalue and eigenvector patterns, scaling-rotation principal geodesic analysis, and a scaling-rotation regresssion framework. The scaling-rotation framework shold also be particularly amenable to diffusion tensor processing since the eigenvectors and eigenvalues of a diffusion tensor model the principal directions of water diffusion and diffusion intensity at a particular voxel.

We recommend using the scaling-rotation estimatio procedure presented in this chapter for $p = 2, 3$ since the number of eigen-decompositions of an SPD matrix from S_p^{top} grows rapidly with p, and procedures for determining a minimal pair of eigen-decompositions for $p > 3$ for arbitrary eigenvalue multiplicity patterns do not yet exist. A computational procedure for calculating the sample minimal scaling-rotation mean set can be developed for $p = 2, 3$, and establishing asymptotic properties for the sample minimal scaling-rotation set could provide interesting future work.

We also note that in subsection 3.2.2, we have established conditions under which our procedure for estimating the sample partial minimal scaling-rotation mean will yield a solution which is unique up to the group action of $\mathcal{G}(p)$, however, we have not yet established

conditions under which the sample partial minimal scaling-rotation mean, as defined in definition 3.1.4, is unique up to the group action of $\mathcal{G}(p)$. We hypothesize that this result will hold if the observations $S_1, \ldots, S_n$ are sufficiently concentrated, however we have not yet established this kind of result. It would also be interesting to establish a similar result for the population partial minimal scaling-rotation mean.

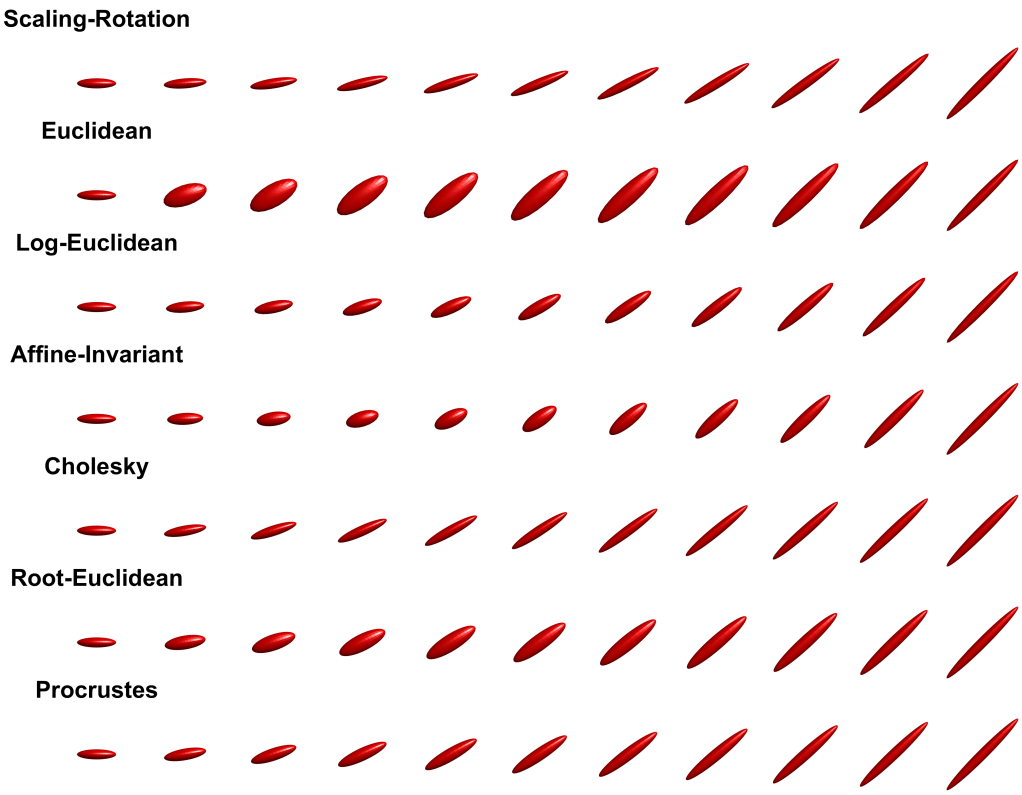

Figure 8: Seven different geodesic paths between X_1 and Y_1.

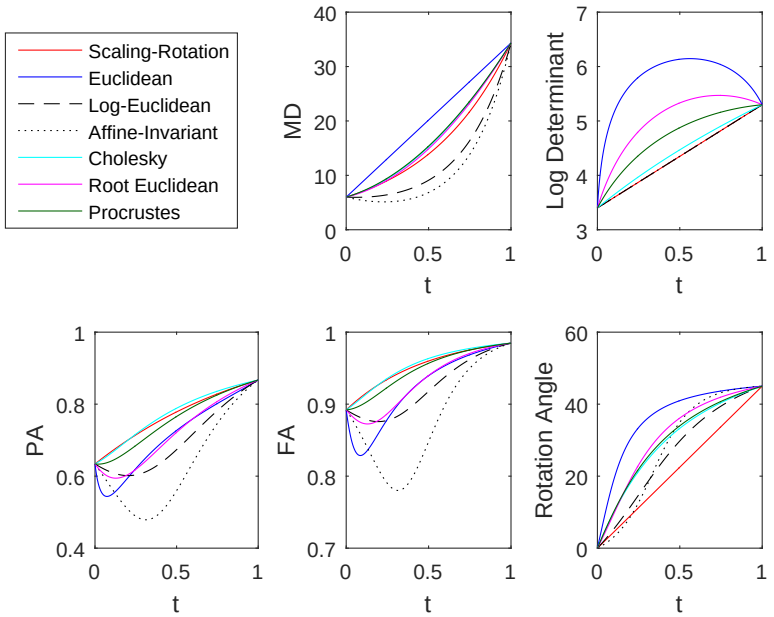

Figure 9: DTI summary measures along each geodesic path connecting X_1 and Y_1.

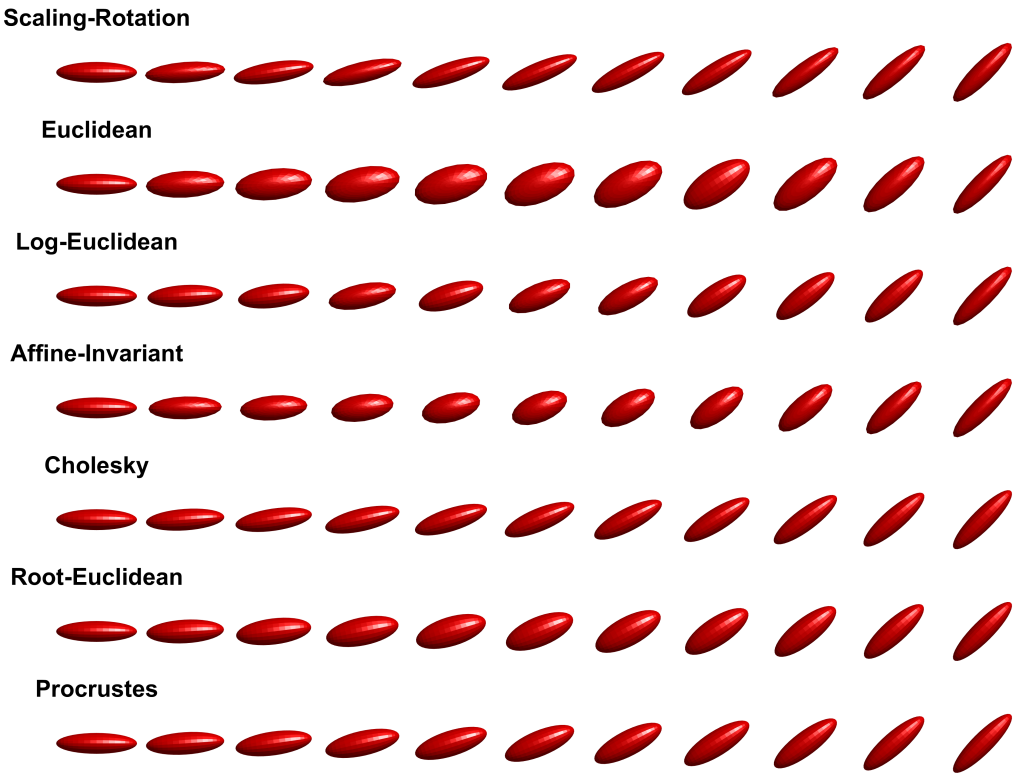

Figure 10: Seven different geodesic paths between X_2 and Y_2.

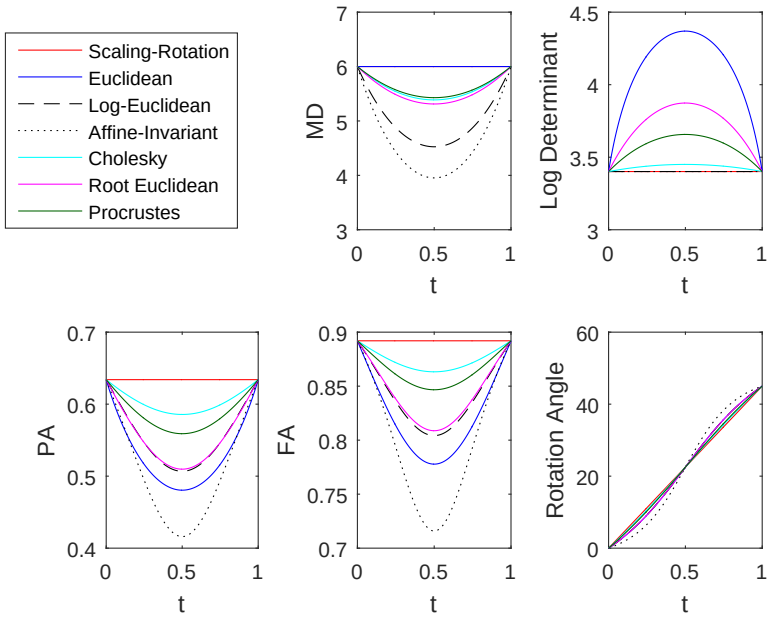

Figure 11: DTI summary measures along each geodesic path connecting X_2 and Y_2.

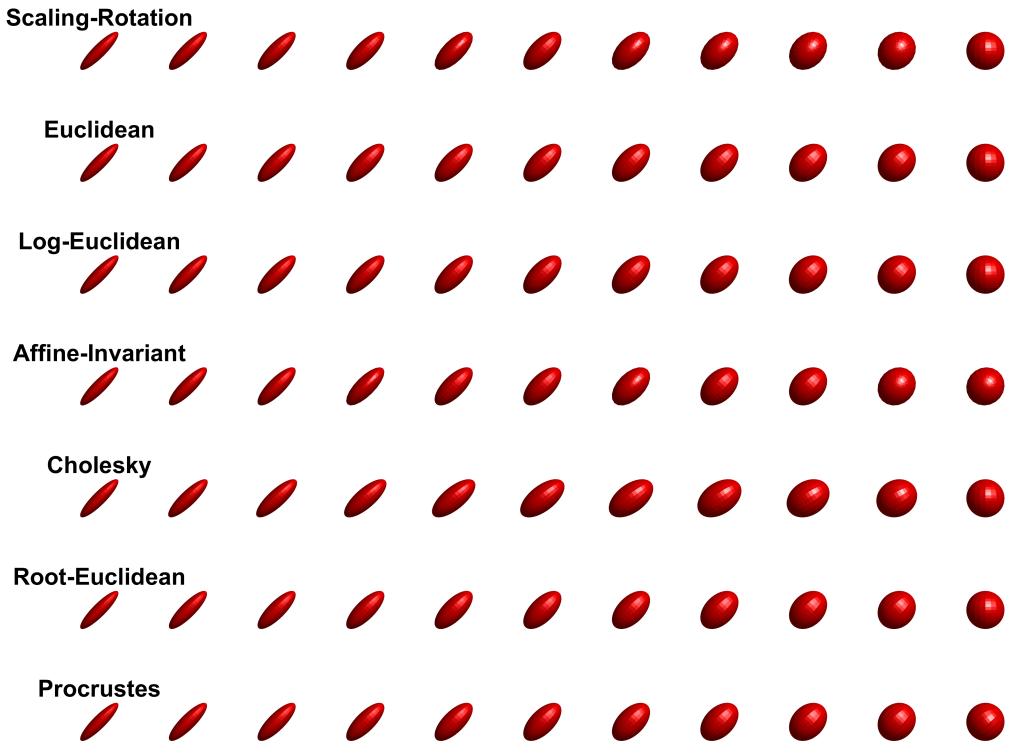

Figure 12: Seven different geodesic paths between X_3 and Y_3.

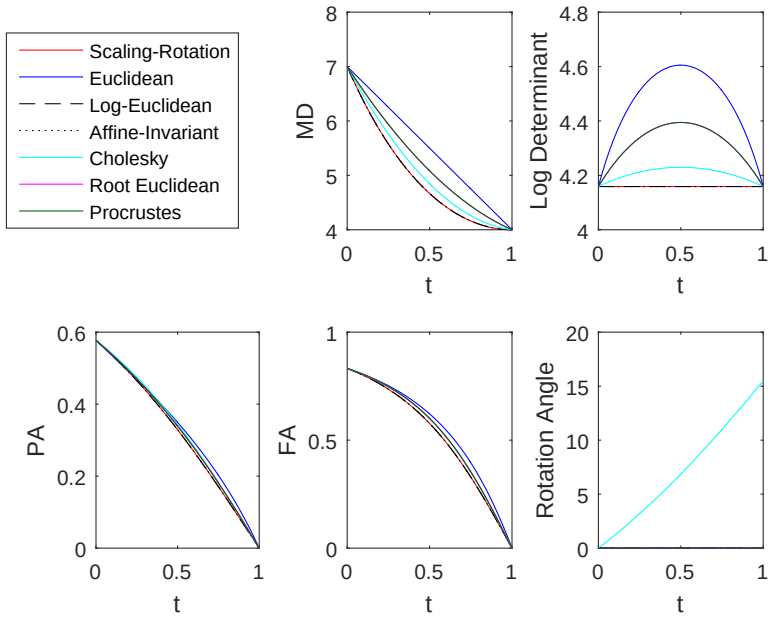

Figure 13: DTI summary measures along each geodesic path connecting X_3 and Y_3.

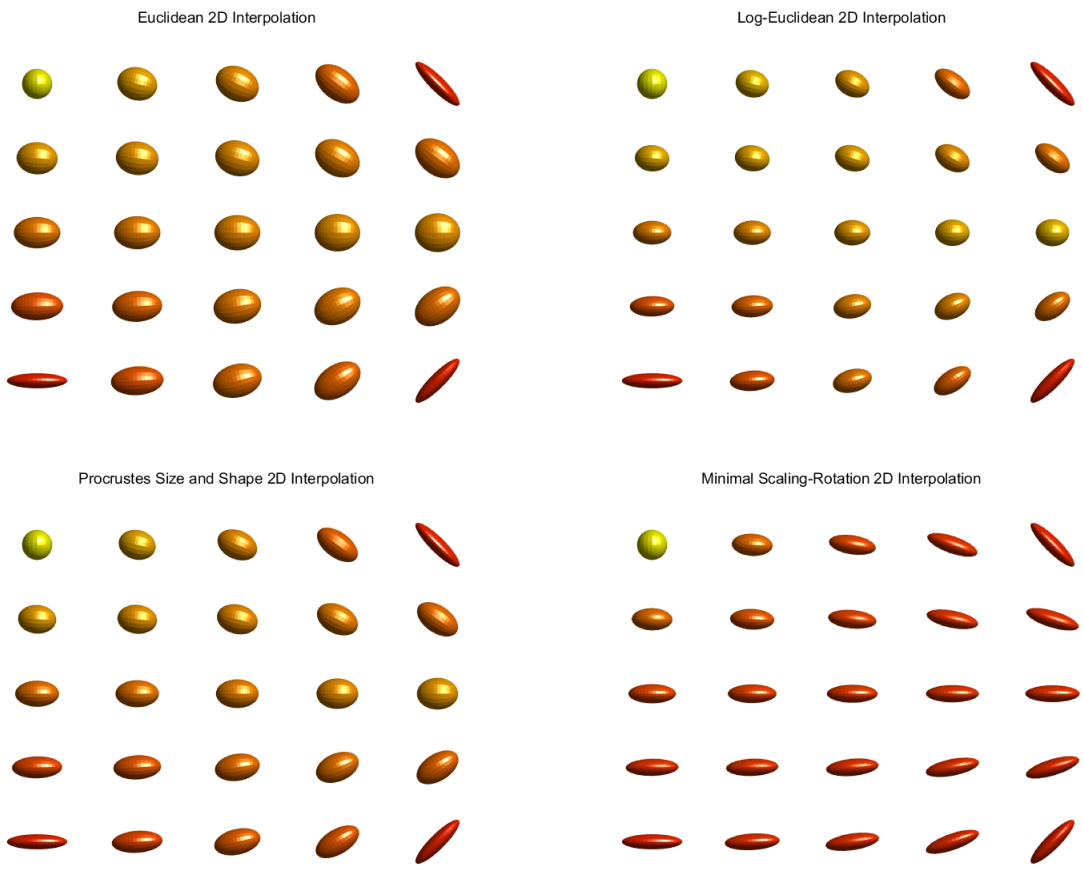

Figure 14: 2D interpolation of 4 diffusion tensors.

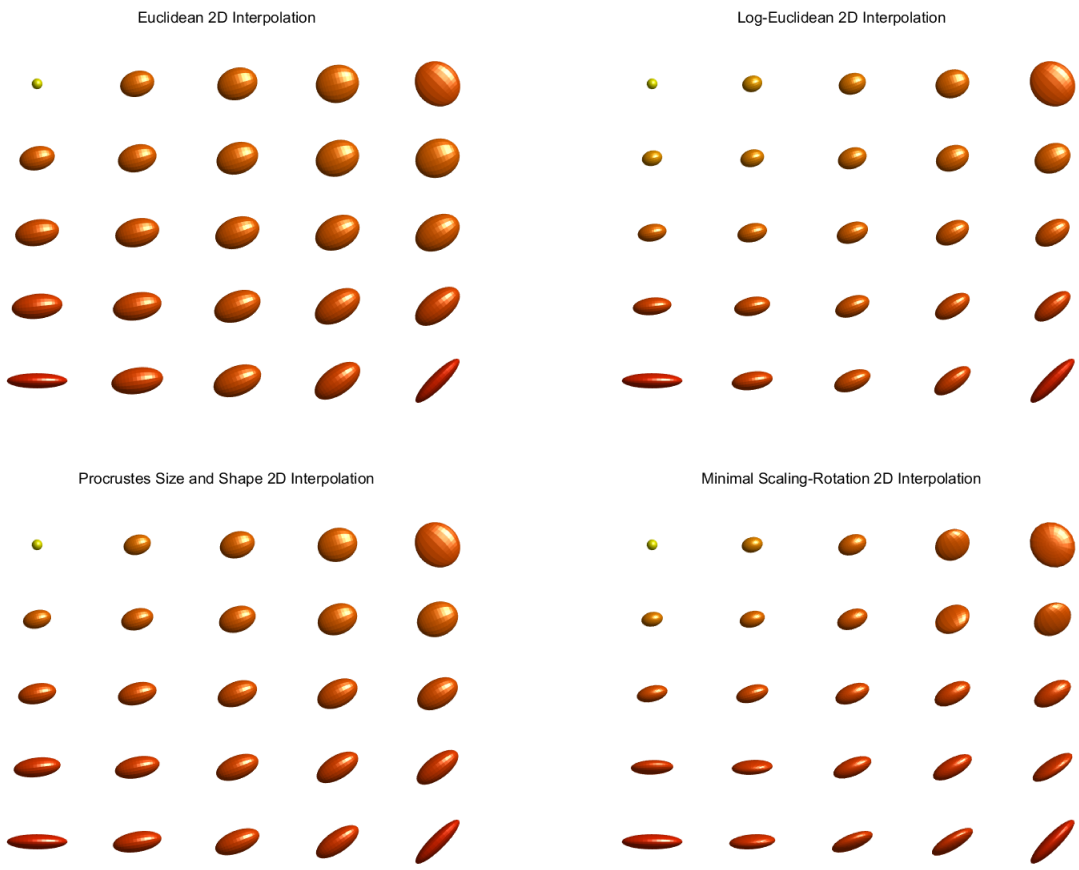

Figure 15: 2D interpolation of 4 diffusion tensors.

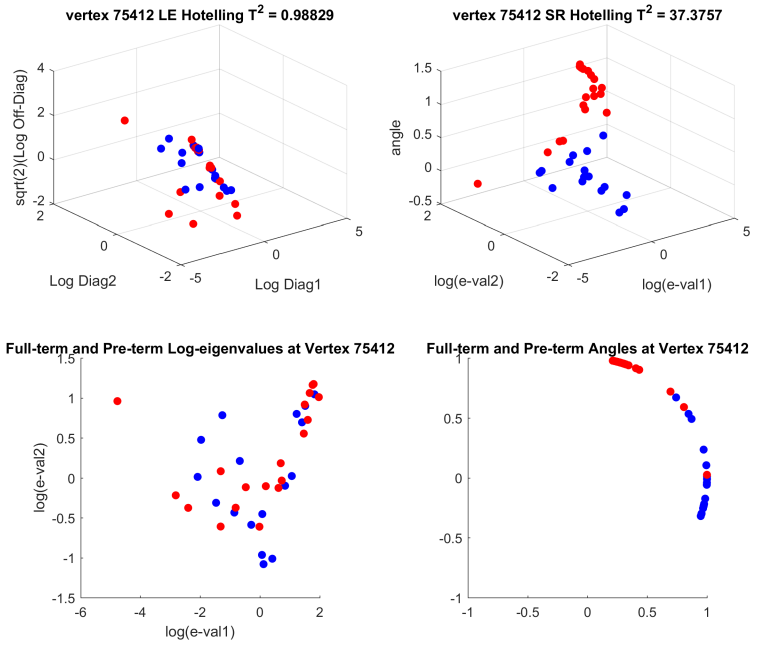

Figure 16: Observations from preterm infants are blue and observations from fullterm infants are red. (Top Left) Linearized deformation tensors used to compute the Log-Euclidean Hotelling's T^2 test statistic. (Top Right) Linearized deformation tensors used to compute the scaling-rotation Hotelling's T^2 test statistic. (Bottom Left) Plot of log-eigenvalues from Top Right plot. (Bottom Right) Plot of angles from Top Right plot mapped to the unit circle in $\mathbb{R}^2$.

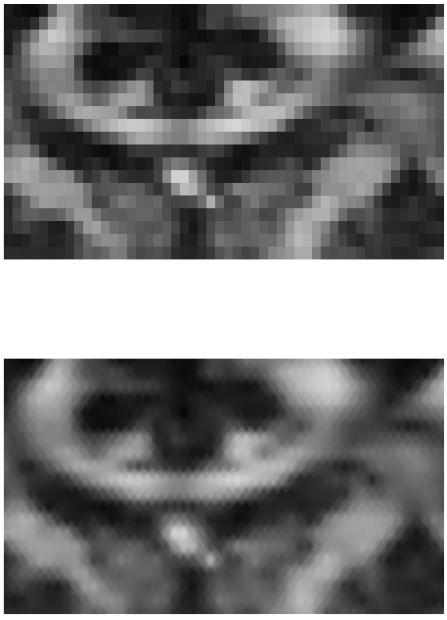

Figure 17: (Top) Fractional Anisotropy map of the original DTI data. (Bottom) Fractional Anisotropy map of interpolated DTI data.

4.0 MOMENT BASED ESTIMATION AND INFERENCE OF EIGENVALUE PARAMETERS

In the visualization of diffusion tensor imaging data, it is often of interest to describe and quantify the shape of diffusion tensors via anisotropy measures such as fractional anisotropy, which was defined in Equation 3.12. Anisotropy describes the spread of the eigenvalues of a diffusion tensor, with emphasis on how distinct the largest, or principal, eigenvalue is from the two smaller eigenvalues. When the spread of the eigenvalues is very small, a diffusion tensor will look almost spherical, and when the largest eigenvalue is much larger than the two smaller eigenvalues, the diffusion tensor will have one very long axis corresponding to the principal eigenvalue and two comparatively short axes corresponding to the two smaller eigenvalues. These two cases, visualized in Figure 7, are known as isotropy and anisotropy, respectively.

The shape of a diffusion tensor with eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ can also be described by its eigenvalue multiplicity pattern. There are four possible eigenvalue multiplicity patterns which correspond to four different diffusion tensor shapes:

- (Isotropic) $\lambda_1 = \lambda_2 = \lambda_3$
- (Prolate) $\lambda_1 > \lambda_2 = \lambda_3$
- (Oblate) $\lambda_1 = \lambda_2 > \lambda_3$
- (Triaxial) $\lambda_1 > \lambda_2 > \lambda_3$

Isotropic diffusion tensors are spherical, prolate diffusion tensors are football-shaped, and oblate diffusion tensors resemble discs.

In this chapter we consider the situation in which we observe a random sample of diffusion tensors $S_1, \ldots, S_n \in Sym^+(3)$ and we would like to estimate and perform inference

about the multiplicity pattern of the population eigenvalue mean. In [44], the authors derive hypothesis tests concerning the multiplicity pattern of the population eigenvalue mean assuming that the data follows a specific type of normal distribution for SPD matrices. In this chapter, we derive estimation procedures and hypothesis tests about multiplicity patterns of the population mean of the log-eigenvalue distribution of diffusion tensors making no distributional assumptions except for finite second moments and an isotropic covariance structure of the log-eigenvalues.

Suppose that we observe a random sample of SPD matrices $S_1, \ldots, S_n$. We assume that each SPD matrix arises from the same generating distribution on $Diag^+(p) \times SO(p)$. Denote the distribution of the vectorized log-eigenvalues from the generating distribution as G with the following assumptions

1. G is a continuous distribution on $\mathbb{R}^p$.
2. G has an unknown mean $\boldsymbol{\mu} \in \mathbb{R}^p$.
3. G has an isotropic covariance structure with unknown variance $\sigma^2 > 0$.

Since G is continuous, each SPD matrix will have no repeated eigenvalues with probability 1, and will have $2^{p-1}p!$ distinct eigen-decompositions. Let $(D_1, U_1), \ldots, (D_n, U_n)$ denote the unobservable generating eigen-decompositions of $S_1, \ldots, S_n$. The log-eigenvalue vectors $\mathbf{X}_1 = vec(Log(D_1)), \ldots, \mathbf{X}_n = vec(Log(D_n))$ are i.i.d. observations from $G(\boldsymbol{\mu}, \sigma^2)$.

To estimate $\boldsymbol{\mu}$, we will first pick observed eigen-decompositions $(D_1', U_1'), \ldots, (D_n', U_n')$ as random draws from the uniform distributions on $\mathcal{F}^{-1}(S_1), \ldots, \mathcal{F}^{-1}(S_n)$. The vectorized log-eigenvalues from observed eigen-decomposition (D_i', U_i') have the representation

$$\mathbf{Y}_i = vec(Log(D_i')) = Z_i \mathbf{X}_i, \tag{4.1}$$

where $\mathbf{X}_i \sim G(\boldsymbol{\mu}, \sigma^2)$ and Z_i is a permuation matrix randomly drawn from the uniform distribution on the set of $p \times p$ permutation matrices.

From the assumptions on $G(\boldsymbol{\mu}, \sigma^2)$ it can be shown that for $\mathbf{Y} = Z\mathbf{X} = (Y_1, \ldots, Y_p)^T$, where $\mathbf{X} \sim G(\boldsymbol{\mu}, \sigma^2)$ and Z follows the uniform distribution on the $p \times p$ permutation

matrices, the following moment equations hold:

$$\alpha_1 = E[Y_1] = \cdots = E[Y_p] = \frac{1}{p}\sum_{i=1}^{p}\mu_i. \tag{4.2}$$

$$\alpha_2 = E[Y_1^2] = \cdots = E[Y_p^2] = \sigma^2 + \frac{1}{p}\sum_{i=1}^{p}\mu_i^2. \tag{4.3}$$

$$\alpha_3 = E[Y_k Y_\ell] = \frac{1}{\binom{p}{2}}\sum_{1\le i<j\le p}\mu_i\mu_j \text{ for any } k \neq \ell. \tag{4.4}$$

$$\alpha_4 = E[Y_h Y_\ell Y_m] = \frac{1}{\binom{p}{3}}\sum_{1\le i<j<k\le p}\mu_i\mu_j\mu_k \text{ for } h \neq \ell \neq m \text{ and } p \geq 3. \tag{4.5}$$

We will use the moment equations above to develop moment-based hypothesis tests of multplicity patterns for $\boldsymbol{\mu}$.

4.1 MOMENT-BASED HYPOTHESIS TEST FOR ISOTROPIC EIGENVALUE MEAN

We first present a test of the null hypothesis $H_0 : \mu_1 = \cdots = \mu_p$ versus the alternative hypothesis $H_A : \mu_i \neq \mu_j$ for some $i \neq j$. From (4.2) and (4.4), it can be shown that

$$g(\alpha_1, \alpha_3) = \alpha_1^2 - \alpha_3 = \frac{1}{p^2(p-1)}\sum_{1\le i<j\le p}(\mu_i - \mu_j)^2,$$

which equals 0 if and only if $\mu_1 = \cdots = \mu_p$.

Using sample estimators $\hat{\alpha}_1$ and $\hat{\alpha}_3$ for α_1 and α_3, we will create a test statistic for the hypotheses above based on the quantity $g(\hat{\alpha}_1, \hat{\alpha}_3)$. Given a random sample of vectorized log-eigenvalues $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ with $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^T$, we define the sample estimators $\hat{\alpha}_1$ and $\hat{\alpha}_3$ as

$$\hat{\alpha}_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{p}\sum_{j=1}^{p}Y_{ij}\right) = \frac{1}{n}\sum_{i=1}^{n}W_i \tag{4.6}$$

$$\hat{\alpha}_3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\binom{p}{2}}\sum_{1\le j<k\le p}Y_{ij}Y_{ik}\right) = \frac{1}{n}\sum_{i=1}^{n}V_i. \tag{4.7}$$

By the Central Limit Theorem, $\sqrt{n}(\hat{\alpha}_1-\alpha_1, \hat{\alpha}_3-\alpha_3)^T \to N_2(0,\Gamma)$. Note that $g(x,y) = x^2-y$ has gradient at (α_1, α_3) equal to

$$\nabla g(\alpha_1,\alpha_3) = \left(\frac{\partial g}{\partial x}\Big|_{\substack{(x,y)= \\ (\alpha_1,\alpha_3)}} = 2\alpha_1, \frac{\partial g}{\partial y}\Big|_{\substack{(x,y)= \\ (\alpha_1,\alpha_3)}} = -1 \right)^T,$$

which does not equal $(0,0)$ for any combination of α_1 and α_3 produced by μ. From an application of the Delta Method, it follows that

$$\sqrt{n}(g(\hat{\alpha}_1,\hat{\alpha}_3) - g(\alpha_1,\alpha_3)) \xrightarrow{L} N(0, \nabla g(\alpha_1,\alpha_3)^T \Gamma \nabla g(\alpha_1,\alpha_3)).$$

We will estimate the asymptotic variance in the equation above with the consistent estimator $\nabla g(\hat{\alpha}_1,\hat{\alpha}_3)^T \hat{\Gamma} \nabla g(\hat{\alpha}_1,\hat{\alpha}_3)$, where

$$\begin{aligned}
\nabla g(\hat{\alpha}_1,\hat{\alpha}_3) &= (2\hat{\alpha}_1, -1)^T \\
\hat{\Gamma}_{11} &= \frac{1}{n}\sum_{i=1}^{n}(W_i - \hat{\alpha}_1)^2 \\
\hat{\Gamma}_{12} = \hat{\Gamma}_{21} &= \frac{1}{n}\sum_{i=1}^{n}(W_i - \hat{\alpha}_1)(V_i - \hat{\alpha}_3) \\
\hat{\Gamma}_{22} &= \frac{1}{n}\sum_{i=1}^{n}(V_i - \hat{\alpha}_3)^2.
\end{aligned}$$

Under the null hypothesis, it follows that

$$T_n = \frac{\sqrt{n} g(\hat{\alpha}_1,\hat{\alpha}_3)}{\sqrt{\nabla g(\hat{\alpha}_1,\hat{\alpha}_3)^T \hat{\Gamma} \nabla g(\hat{\alpha}_1,\hat{\alpha}_3)}} \xrightarrow{L} N(0,1) \tag{4.8}$$

by Slutsky's Theorem. To create an asymptotic α-level test of H_0 vs. H_A, reject H_0 if $T_n > \Phi^{-1}(1-\alpha)$ where $\Phi^{-1}(.)$ denotes the inverse cdf of the standard normal distribution.

We are using a one-sided test of $H_0 : \mu_1 = \cdots = \mu_p$ because if the alternative hypothesis is true, for large n, the test statistic T_n will approximately follow a normal distribution with positive mean $\sqrt{n} g(\alpha_1,\alpha_3)/\sqrt{\nabla g(\alpha_1,\alpha_3)^T \Gamma \nabla g(\alpha_1,\alpha_3)}$ and variance 1. Thus, having a test statistic T_n that is large and positive should provide evidence in favor of the alternative hypothesis.

To illustrate the convergence of the test statistic T_n to a standard normal distribution under the null hypothesis that $\mu_1 = \cdots = \mu_p$, we have performed a simulation experiment

drawing 1000 samples of size $n = 50, 100,$ and 500 from the bivariate normal distribution with mean $\mu = (2, 2)^T$ and covariance matrix $0.75I_2$. To simulate each observation as the log-eigenvalues from a randomly chosen eigen-decomposition, the components of each observation were permuted with probability $1/2$. For each sample of observed log-eigenvalues, we computed the test statistic T_n using the formula from (4.8), resulting in 1000 test statistics for each of the sample sizes 50, 100, and 500. In Figure 18, we can see that the distribution of T_n is very similar to a standard normal distribution with $n = 50$ and appears even closer to normal for larger sample sizes. As can be seen in Table 1, the empirical type-I error rate is generally close to the intended significance level, and not surprisingly, the empirical error rate is closest to the intended level at the largest sample size of $n = 500$, suggesting that the asymptotic type-I error rate of the hypothesis test will equal the intended significance level of the test.

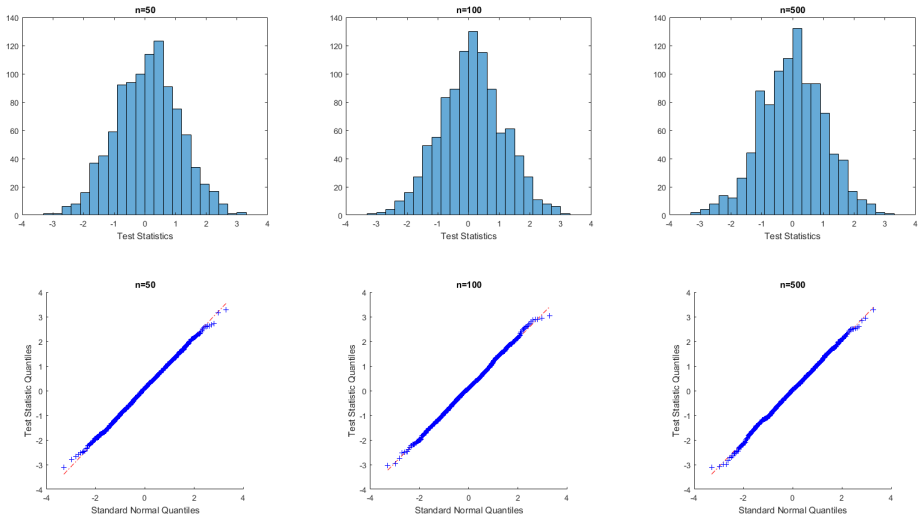

Figure 18: Histograms and quantile-quantile plots to illustrate the convergence of T_n to $N(0, 1)$ in distribution under $H_0 : \mu_1 = \mu_2$.

Table 1: Empirical Type I Error Rates

Asym. α	$n = 50$	$n = 100$	$n = 500$
0.01	0.0130	0.0180	0.0120
0.05	0.0630	0.0710	0.0590
0.10	0.1240	0.1380	0.1100

4.2 METHOD OF MOMENTS ESTIMATION FOR P=2

The derivation of the moment-based hypothesis test of an isotropic log-eigenvalue mean can also lead to moment-based estimators for μ when $p = 2$. It can be shown that

$$g(\alpha_1, \alpha_3) = \alpha_1^2 - \alpha_3 = \left(\frac{\mu_1 - \mu_2}{2}\right)^2$$

when $p = 2$. Then $\mu_{(1)} = \max\{\mu_1, \mu_2\}$ and $\mu_{(2)} = \min\{\mu_1, \mu_2\}$ can be recovered from the following equations

$$\mu_{(1)}(\alpha_1, \alpha_3) = \alpha_1 + \sqrt{\alpha_1^2 - \alpha_3}$$
$$\mu_{(2)}(\alpha_1, \alpha_3) = \alpha_1 - \sqrt{\alpha_1^2 - \alpha_3}.$$

Note that the mean μ can only be estimated up to permutations of its entries because, if μ is unknown then we will not know if $\mu = (\mu_{(1)}, \mu_{(2)})^T$ or $\mu = (\mu_{(2)}, \mu_{(1)})^T$.

Given $Y_1, \ldots, Y_n$, a random sample of vectorized log-eigenvalues from observed eigendecompositions, we begin with $\mu^{desc}(\hat{\alpha}_1, \hat{\alpha}_3) = (\mu_{(1)}(\hat{\alpha}_1, \hat{\alpha}_3), \mu_{(2)}(\hat{\alpha}_1, \hat{\alpha}_3))^T$ as a sample estimator for $\mu^{desc}(\alpha_1, \alpha_3) = (\mu_{(1)}(\alpha_1, \alpha_3), \mu_{(2)}(\alpha_1, \alpha_3))^T$, where $\hat{\alpha}_1$ and $\hat{\alpha}_3$ are defined as in (4.6) and (4.7). The quantity $\hat{\alpha}_1^2 - \hat{\alpha}_3$ can be negative, so to avoid obtaining complex-valued estimates for $\mu^{desc}(\alpha_1, \alpha_3)$, we will make the following modification

$$\mu^{desc}(\hat{\alpha}_1, \hat{\alpha}_3) = \begin{cases} (\mu_{(1)}(\hat{\alpha}_1, \hat{\alpha}_3), \mu_{(2)}(\hat{\alpha}_1, \hat{\alpha}_3))^T & \text{if } \hat{\alpha}_1^2 - \hat{\alpha}_3 \geq 0 \\ (\hat{\alpha}_1, \hat{\alpha}_1)^T & \text{if } \hat{\alpha}_1^2 - \hat{\alpha}_3 < 0 \end{cases}.$$

Table 2: Empirical Mean Squared Error (MSE) from Simulations of $\mu^{desc}(\hat{\alpha}_1, \hat{\alpha}_3)$

$\boldsymbol{\mu}$	$n = 50$	$n = 100$	$n = 500$
$\mu_1 = \mu_2$	0.1093	0.0756	0.0291
$\mu_1 \neq \mu_2$	0.0390	0.0191	0.0036

Alternatively, the conditions above could depend on the test statistic T_n from the previous subsection since the sign of T_n is determined by $\hat{\alpha}_1^2 - \hat{\alpha}_3$. Thus, we could also write $\mu^{desc}(\hat{\alpha}_1, \hat{\alpha}_3)$ in a more compact form

$$\mu^{desc}(\hat{\alpha}_1, \hat{\alpha}_3) = (\hat{\alpha}_1, \hat{\alpha}_1)^T + I\{T_n \geq 0\}\sqrt{\hat{\alpha}_1^2 - \hat{\alpha}_3}(1, -1)^T, \tag{4.9}$$

where $I\{T_n \geq 0\}$ denotes the indicator function for the event "$T_n \geq 0$".

The estimator $\mu^{desc}(\hat{\alpha}_1, \hat{\alpha}_3)$ will be a consistent estimator of $\mu^{desc}(\alpha_1, \alpha_3)$. Regardless of the multiplicity pattern of μ, the estimators $\hat{\alpha}_1$ and $\hat{\alpha}_3$ will be consistent estimators of α_1 and α_3 by the Law of Large Numbers, which implies that $\sqrt{\hat{\alpha}_1^2 - \hat{\alpha}_3}$ will be a consistent estimator of $\sqrt{\alpha_1^2 - \alpha_3}$. If $\mu_1 = \mu_2$, then $\mu^{desc}(\hat{\alpha}_1, \hat{\alpha}_3) \xrightarrow{p} (\alpha_1, \alpha_1)^T = \mu^{desc}(\alpha_1, \alpha_3)$ because $\sqrt{\alpha_1^2 - \alpha_3} \xrightarrow{p} \sqrt{\alpha_1^2 - \alpha_3} = 0$ and $T_n \xrightarrow{L} N(0, 1)$, which implies that $I\{T_n \geq 0\}$ will converge to a Bernoulli random variable with probability of success $1/2$. If $\mu_1 \neq \mu_2$, then T_n will have the same asymptotic behavior as $Z_n = Z + \sqrt{n}g(\alpha_1, \alpha_3)/\sqrt{\nabla g(\alpha_1, \alpha_3)^T \Gamma \nabla g(\alpha_1, \alpha_3)}$, where $Z \sim N(0, 1)$. Then we have that $P(T_n \geq 0) \to 1 - \Phi(-\infty) = 1$, which implies that $\mu^{desc}(\hat{\alpha}_1, \hat{\alpha}_3) \xrightarrow{p} \mu^{desc}(\alpha_1, \alpha_3)$.

To illustrate the consistency of $\mu^{desc}(\hat{\alpha}_1, \hat{\alpha}_3)$, we have performed a simulation experiment drawing 1000 samples of size $n = 50, 100,$ and 500 from bivariate normal distributions $N_2(\boldsymbol{\mu}_1, \sigma^2 I_2)$ and $N_2(\boldsymbol{\mu}_2, \sigma^2 I_2)$ with mean $\boldsymbol{\mu}_1 = (2, 2)^T$, $\boldsymbol{\mu}_2 = (2 + \sqrt{3/4}, 2 - \sqrt{3/4})^T$, and $\sigma^2 = 3/4$. To simulate each observation as the log-eigenvalues from a randomly chosen eigendecomposition, the components of each observation were permuted with probability $1/2$. For each sample of observed log-eigenvalues, we computed $\mu^{desc}(\hat{\alpha}_1, \hat{\alpha}_3)$ using the formula from (4.9), resulting in 1000 estimates for each combination of means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ with the sample sizes 50, 100, and 500. As can be seen in Table 2, the empirical mean-squared error (MSE) decreases as the sample size increases under both assumptions about the mutliplicity pattern of $\boldsymbol{\mu}$, supporting the consistency arguments above.

4.3 MOMENT-BASED HYPOTHESIS TEST FOR EIGENVALUE MEAN IN MIDDLE OR LOWER STRATUM (P=3)

We next present a hypothesis test of $H_0 : \mu$ has at most 2 unique values versus the alternative $H_A : \mu$ has three unique values. From the moment equations for α_1 and α_3 when $p = 3$, it can be shown that

$$\mu_3 = 3\alpha_1 - (\mu_1 + \mu_2) \tag{4.10}$$

$$\mu_1^2 + \mu_2^2 + \mu_3^2 = 3(3\alpha_1^2 - 2\alpha_3). \tag{4.11}$$

Plugging (4.10) into (4.11) yields

$$\mu_1^2 + \mu_2^2 + \mu_1\mu_2 - 3\alpha_1(\mu_1 + \mu_2) + 3\alpha_3 = 0, \tag{4.12}$$

which is the equation for an ellipse in (μ_1, μ_2) that is centered at (α_1, α_1). Equation 4.12 can be simplified as $v^T A v = 3(\alpha_1^2 - \alpha_3)$ where

$$v = (\mu_1 - \alpha_1, \mu_2 - \alpha_1)^T, A = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}.$$

From this elliptical equation and (4.10), we can solve for the coordinates of $\mu_1, \mu_2,$ and μ_3 as

$$\mu_1 = \alpha_1 + \sqrt{3(\alpha_1^2 - \alpha_3)}\left(-\cos(\theta) + \frac{1}{\sqrt{3}}\sin(\theta)\right) \tag{4.13}$$

$$\mu_2 = \alpha_1 + \sqrt{3(\alpha_1^2 - \alpha_3)}\left(\cos(\theta) + \frac{1}{\sqrt{3}}\sin(\theta)\right) \tag{4.14}$$

$$\mu_3 = \alpha_1 - 2\sqrt{(\alpha_1^2 - \alpha_3)}\sin(\theta). \tag{4.15}$$

Plugging these coordinates for $\mu_1, \mu_2,$ and μ_3 into the moment equation for α_4 when $p = 3$ yields the equality

$$\alpha_4 - \alpha_1^3 + 3\alpha_1(\alpha_1^2 - \alpha_3) = 2(\alpha_1^2 - \alpha_3)^{3/2}\sin(3\theta). \tag{4.16}$$

From (4.16), we will show that the quantity

$$h(\alpha_1, \alpha_3, \alpha_4) = 4(\alpha_1^2 - \alpha_3)^3 - (\alpha_4 - \alpha_1^3 + 3\alpha_1(\alpha_1^2 - \alpha_3))^2$$

equals 0 if and only if μ has at most 2 unique values. When $\mu_1 = \mu_2 = \mu_3$, it follows that $\alpha_4 - \alpha_1^3 = 0$ and $\alpha_1^2 - \alpha_3 = 0$, which imply that $h(\alpha_1, \alpha_3, \alpha_4) = 0$. If two of the entries of μ are equal, one of the cases below must be true:

$$\mu_1 - \mu_2 = -2\sqrt{3(\alpha_1^2 - \alpha_3)}\cos(\theta) = 0 \iff \cos(\theta) = 0 \iff \theta \in \left\{\frac{\pi}{2}, \frac{3\pi}{2}\right\} \tag{4.17}$$

$$\mu_2 - \mu_3 = \cos(\theta) + \sqrt{3}\sin(\theta) = 0 \iff \frac{\sin(\theta)}{\cos(\theta)} = -\frac{1}{\sqrt{3}} \iff \theta \in \left\{\frac{5\pi}{6}, \frac{11\pi}{6}\right\} \tag{4.18}$$

$$\mu_1 - \mu_3 = -\cos(\theta) + \sqrt{3}\sin(\theta) = 0 \iff \frac{\sin(\theta)}{\cos(\theta)} = \frac{1}{\sqrt{3}} \iff \theta \in \left\{\frac{\pi}{6}, \frac{7\pi}{6}\right\}. \tag{4.19}$$

For $\theta \in \{\frac{\pi}{6}, \frac{5\pi}{6}, \frac{3\pi}{2}\}$, $\sin(3\theta) = 1$, while $\sin(3\theta) = -1$ for $\theta \in \{\frac{\pi}{2}, \frac{7\pi}{6}, \frac{11\pi}{6}\}$. For all of these cases, it follows from (4.16) that $|\alpha_4 - \alpha_1^3 + 3\alpha_1(\alpha_1^2 - \alpha_3)| = |2(\alpha_1^2 - \alpha_3)^{3/2}|$, which implies that $h(\alpha_1, \alpha_3, \alpha_4) = 0$.

If all three entries of μ are distinct, then we will have that $\theta \notin \{\frac{\pi}{6}, \frac{5\pi}{6}, \frac{3\pi}{2}, \frac{\pi}{2}, \frac{7\pi}{6}, \frac{11\pi}{6}\}$, which, from (4.16), implies that $|\alpha_4 - \alpha_1^3 + 3\alpha_1(\alpha_1^2 - \alpha_3)| < |2(\alpha_1^2 - \alpha_3)^{3/2}|$ and $h(\alpha_1, \alpha_3, \alpha_4) > 0$. To summarize, the quantity $h(\alpha_1, \alpha_3, \alpha_4)$ will equal 0 when μ has at most 2 unique entries, and will be positive when μ has three unique entries.

Using sample estimators $\hat{\alpha}_1$, $\hat{\alpha}_3$, and $\hat{\alpha}_4$ for the population moments α_1, α_3, and α_4, we will create a test statistic for the null hypothesis that μ has at most 2 distinct entries based on the sample statistic $h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)$. Given a random sample of vectorized log-eigenvalues $Y_1, \ldots, Y_n$ with $Y_i = (Y_{i1}, Y_{i2}, Y_{i3})^T$, we define the sample estimators $\hat{\alpha}_1$, $\hat{\alpha}_3$, and $\hat{\alpha}_4$ as

$$\hat{\alpha}_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{3}\sum_{j=1}^{3} Y_{ij}\right) = \frac{1}{n}\sum_{i=1}^{n} T_i \tag{4.20}$$

$$\hat{\alpha}_3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\binom{3}{2}}\sum_{1\le j<k\le 3} Y_{ij}Y_{ik}\right) = \frac{1}{n}\sum_{i=1}^{n} W_i \tag{4.21}$$

$$\hat{\alpha}_4 = \frac{1}{n}\sum_{i=1}^{n} Y_{i1}Y_{i2}Y_{i3}. \tag{4.22}$$

By the Central Limit Theorem, $\sqrt{n}(\hat{\alpha}_1 - \alpha_1, \hat{\alpha}_3 - \alpha_3, \hat{\alpha}_4 - \alpha_4) \xrightarrow{L} N_3(0, \Gamma)$. The gradient of the function $h(x, y, z)$ evaluated at $(\alpha_1, \alpha_3, \alpha_4)$, which we will denote as $\nabla h(\alpha_1, \alpha_3, \alpha_4)$,

will have partial derivatives

$$\frac{\partial h}{\partial x}\Big|_{\substack{(x,y,z)=\\(\alpha_1,\alpha_3,\alpha_4)}} = 24\alpha_1(\alpha_1^2-\alpha_3)^2 - 2(\alpha_4-\alpha_1^3+3\alpha_1(\alpha_1^2-\alpha_3))(3\alpha_1^2+3(\alpha_1^2-\alpha_3)) \tag{4.23}$$

$$\frac{\partial h}{\partial y}\Big|_{\substack{(x,y,z)=\\(\alpha_1,\alpha_3,\alpha_4)}} = 6\alpha_1(\alpha_4-\alpha_1^3+3\alpha_1(\alpha_1^2-\alpha_3)) - 12(\alpha_1^2-\alpha_3)^2 \tag{4.24}$$

$$\frac{\partial h}{\partial z}\Big|_{\substack{(x,y,z)=\\(\alpha_1,\alpha_3,\alpha_4)}} = -2(\alpha_4-\alpha_1^3+3\alpha_1(\alpha_1^2-\alpha_3)). \tag{4.25}$$

If $\mu_1=\mu_2=\mu_3$, then $\alpha_4-\alpha_1^3=0$ and $\alpha_1^2-\alpha_3=0$, which imply that the gradient equals $(0,0,0)^T$. If μ only has two distinct entries, then $\alpha_1^2-\alpha_3>0$ and $|\alpha_4-\alpha_1^3+3\alpha_1(\alpha_1^2-\alpha_3)| = |2(\alpha_1^2-\alpha_3)^{3/2}|$, which imply that $\frac{\partial h}{\partial z}$ will not equal 0 when evaluated at $(x,y,z)=(\alpha_1,\alpha_3,\alpha_4)$. Thus, $\nabla h(\alpha_1,\alpha_3,\alpha_4)$ will never equal 0 when μ has two distinct entries. We will next show that $\nabla h(\alpha_1,\alpha_3,\alpha_4)$ will never equal 0 when μ has three distinct entries. First consider the case when μ has three distinct entries and $\alpha_4-\alpha_1^3+3\alpha_1(\alpha_1^2-\alpha_3)=0$. In this case, $\frac{\partial h}{\partial y}$ evaluated at $(x,y,z)=(\alpha_1,\alpha_3,\alpha_4)$ will equal $-12(\alpha_1^2-\alpha_3)^2$, which will be non-zero. In the case when μ has three distinct entries and $\alpha_4-\alpha_1^3+3\alpha_1(\alpha_1^2-\alpha_3)\neq 0$, $\frac{\partial h}{\partial z}$ will not equal 0 when evaluated at $(x,y,z)=(\alpha_1,\alpha_3,\alpha_4)$. Thus, we have that $\nabla h(\alpha_1,\alpha_3,\alpha_4)$ only equals 0 when $\mu_1=\mu_2=\mu_3$.

We will next establish the asymptotic behavior of the quantity $\sqrt{n}h(\hat{\alpha}_1,\hat{\alpha}_3,\hat{\alpha}_4)$ under the null hypothesis that μ has at most two unique entries. First suppose that $\mu_1=\mu_2=\mu_3$. Since $\nabla h(\alpha_1,\alpha_3,\alpha_4)$ equals 0, we will use a second order Taylor Expansion to determine the asymptotic distribution of $\sqrt{n}h(\hat{\alpha}_1,\hat{\alpha}_3,\hat{\alpha}_4)$. Since $\sqrt{n}(\hat{\alpha}_1-\alpha_1,\hat{\alpha}_3-\alpha_3,\hat{\alpha}_4-\alpha_4)\xrightarrow{L} N(0,\Gamma)$, $h(\alpha_1,\alpha_3,\alpha_4)=0$, and $\nabla h(\alpha_1,\alpha_3,\alpha_4)=0$, it follows that

$$nh(\hat{\alpha}_1,\hat{\alpha}_3,\hat{\alpha}_4)\xrightarrow{L}\mathbf{Y}^T\nabla^2 h(\alpha_1,\alpha_3,\alpha_4)\mathbf{Y}$$

where $\mathbf{Y}\sim N_3(0,\Gamma)$ and $\nabla^2 h(\alpha_1,\alpha_3,\alpha_4)$ denotes the Hessian matrix of $h(x,y,z)$ evaluated at $(\alpha_1,\alpha_3,\alpha_4)$. When $\mu_1=\mu_2=\mu_3$, it can be shown that

$$\nabla^2 h(\alpha_1,\alpha_3,\alpha_4)=\begin{pmatrix}-18\alpha_1^4 & 18\alpha_1^3 & -6\alpha_1^2\\ 18\alpha_1^3 & -18\alpha_1^2 & 6\alpha_1\\ -6\alpha_1^2 & 6\alpha_1 & -2\end{pmatrix},$$

which is symmetric, has rank 1, and has non-zero eigenvalue $\lambda_H = \text{Tr}(\nabla^2 h(\alpha_1, \alpha_3, \alpha_4)) = (-1)(2 + 18\alpha_1^2 + 18\alpha_1^4)$. If w_H denotes the eigenvector of $\nabla^2 h(\alpha_1, \alpha_3, \alpha_4)$ associated with λ_H, then $\nabla^2 h(\alpha_1, \alpha_3, \alpha_4)$ can be represented as $\nabla^2 h(\alpha_1, \alpha_3, \alpha_4) = \lambda_H w_H w_H^T$. From the eigen-decomposition $\Gamma = UDU^T$ we can represent $\mathbf{Y} \sim N_3(0, \Gamma)$ as $\mathbf{Y} = \Gamma^{1/2}\mathbf{Z} = UD^{1/2}\mathbf{Z}$, where $\mathbf{Z} \sim N_3(0, I_3)$. Then we have that

$$\begin{aligned} nh(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4) &\xrightarrow{L} \mathbf{Y}^T \nabla^2 h(\alpha_1, \alpha_3, \alpha_4)\mathbf{Y} \\ &= \mathbf{Z}^T D^{1/2} U^T \lambda_H w_H w_H^T U D^{1/2}\mathbf{Z} \\ &= \lambda_H \mathbf{Z}^T (D^{1/2}U^T w_H)(D^{1/2}U^T w_H)^T \mathbf{Z}. \end{aligned}$$

Note that the center matrix $(D^{1/2}U^T w_H)(D^{1/2}U^T w_H)^T$ is a rank 1 symmetric matrix with non-zero eigenvalue $\text{Tr}((D^{1/2}U^T w_H)(D^{1/2}U^T w_H)^T) = w_H^T \Gamma w_H$. If v denotes the corresponding eigenvector for this eigenvalue, then we have that

$$nh(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4) \xrightarrow{L} \lambda_H w_H^T \Gamma w_h (v^T \mathbf{Z})^2 = (\lambda_H w_H^T \Gamma w_H)\chi_1^2 \tag{4.26}$$

since $v^T\mathbf{Z} \sim N(0, 1)$. Finally, we have that $\sqrt{n}h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4) = (1/\sqrt{n})nh(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4) \xrightarrow{p} 0$.

Next we establish the asymptotic behavior of $\sqrt{n}h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)$ assuming that μ has two distinct entries. As was shown earlier, $\nabla h(\alpha_1, \alpha_3, \alpha_4)$ does not equal $(0, 0, 0)^T$, so we can apply the Delta Method to get

$$\sqrt{n}h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4) \xrightarrow{L} N(0, \nabla h(\alpha_1, \alpha_3, \alpha_4)^T \Gamma \nabla h(\alpha_1, \alpha_3, \alpha_4))$$

when μ has two distinct entries. We will estimate the asymptotic variance from the equation above with the consistent estimator $\nabla h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)^T \hat{\Gamma} \nabla h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)$, where the entries of Γ

are estimated as

$$\hat{\Gamma}_{11} = \frac{1}{n}\sum_{i=1}^{n}(T_i - \hat{\alpha}_1)^2$$

$$\hat{\Gamma}_{12} = \hat{\Gamma}_{21} = \frac{1}{n}\sum_{i=1}^{n}(T_i - \hat{\alpha}_1)(W_i - \hat{\alpha}_3)$$

$$\hat{\Gamma}_{13} = \hat{\Gamma}_{31} = \frac{1}{n}\sum_{i=1}^{n}(T_i - \hat{\alpha}_1)(Y_{i1}Y_{i2}Y_{i3} - \hat{\alpha}_4)$$

$$\hat{\Gamma}_{22} = \frac{1}{n}\sum_{i=1}^{n}(W_i - \hat{\alpha}_3)^2$$

$$\hat{\Gamma}_{23} = \hat{\Gamma}_{32} = \frac{1}{n}\sum_{i=1}^{n}(W_i - \hat{\alpha}_3)(Y_{i1}Y_{i2}Y_{i3} - \hat{\alpha}_4)$$

$$\hat{\Gamma}_{33} = \frac{1}{n}\sum_{i=1}^{n}(Y_{i1}Y_{i2}Y_{i3} - \hat{\alpha}_4)^2.$$

When μ has two distinct entries, the test statistic

$$V_n = \frac{\sqrt{n}h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)}{\sqrt{\nabla h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)^T \hat{\Gamma} \nabla h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)}} \xrightarrow{L} N(0,1) \tag{4.27}$$

by Slutsky's Theorem. To create an asymptotic α-level test of H_0 vs. H_A, reject H_0 if $V_n > \Phi^{-1}(1-\alpha)$ where $\Phi^{-1}(.)$ denotes the inverse cdf of the standard normal distribution.

We are using a one-sided hypothesis test because if the alternative hypothesis is true, then for large n, the test statistic V_n will approximately follow a normal distribution with positive mean $\sqrt{n}h(\alpha_1, \alpha_3\alpha_4)/\sqrt{\nabla h(\alpha_1, \alpha_3, \alpha_4)^T \Gamma \nabla h(\alpha_1, \alpha_3, \alpha_4)}$ and variance 1. Thus, it makes sense to reject the null hypothesis when U_n is positive and large. Note that when μ has only one unique entry, the test statistic V_n will be negative with very high probability for large n since V_n will approximately follow the distribution of a χ^2_1 random variable scaled by $(\lambda_H w_H^T \Gamma w_H)/\sqrt{n\nabla h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)^T \hat{\Gamma} \nabla h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)} < 0$ for large n (it was previously shown that $\lambda_H = (-1)(2 + 18\alpha_1^2 + 18\alpha_1^4)$).

To illustrate the asymptotic distribution of the test statistic V_n under the null hypothesis, we have performed simulation experiments drawing 1000 samples of size $n = 50, 100, \text{and } 500$ under two cases. Under Case I, samples of size n are drawn from the trivariate normal distribution with mean $\boldsymbol{\mu} = (2,2,2)^T$ and covariance matrix $0.75I_3$. Under Case II, samples of size n are drawn from the trivariate normal distribution with mean

$\boldsymbol{\mu} = (2 + 2\sqrt{3/4}, 2 + 2\sqrt{3/4}, 2)^T$ and covariance matrix $0.75I_3$. To simulate each observation as the log-eigenvalues from a randomly chosen eigen-decomposition, the components of each observation were permuted by a permutation matrix randomly selected from the uniform distribution on the set of 3×3 permutation matrices. For each sample of observed log-eigenvalues, we computed the test statistic V_n using the formula from (4.27), resulting in 1000 test statistics for each of the sample sizes 50, 100, and 500 for each case.

From the simulations under Case I, we see in Figure 19 that the test statistics are negative with high probability and when they are positive, they never exceed $\Phi^{-1}(1-\alpha)$ for $\alpha = 0.01, 0.05, \text{or } 0.10$, as can be seen in Table 3. In Figure 20, we can see the distribution of V_n converging to a standard normal distribution under Case II, and in Table 4, the empirical type I error rate is generally close to the intended value when the sample size is large (greater than 100).

In Figure 21, we illustrate how the speed of convergence of the numerator $\sqrt{n}h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)$ to a normal distribution when $\boldsymbol{\mu}$ has two unique values depends on the separation between its unique values. Let $\mu_{(1)}$ and $\mu_{(2)}$ denote the unique values of $\boldsymbol{\mu}$, with $\mu_{(1)} \geq \mu_{(2)}$. For fixed $n = 100$, we see that the distribution of $\sqrt{n}h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)$ becomes more normal as $\mu_{(1)} - \mu_{(2)}$ increases.

4.4 METHOD OF MOMENTS ESTIMATION FOR P=3

The derivation of the moment-based test that the log-eigenvalue mean has at most two unique entries from the previous subsection can also lead to a moment-based estimator of μ when $p = 3$.

Solving for θ in (4.16) requires locating solutions to

$$\sin(3\theta) = a_0 = \frac{\alpha_4 - \alpha_1^3 + 3\alpha_1(\alpha_1^2 - \alpha_3)}{2(\alpha_1^2 - \alpha_3)^{3/2}}. \tag{4.28}$$

Note that (4.28) cannot be solved when $\mu_1 = \mu_2 = \mu_3$ since the numerator and denominator of a_0 both equal 0 in that case. In this case, $\mu_1 = \mu_2 = \mu_3 = \alpha_1$.

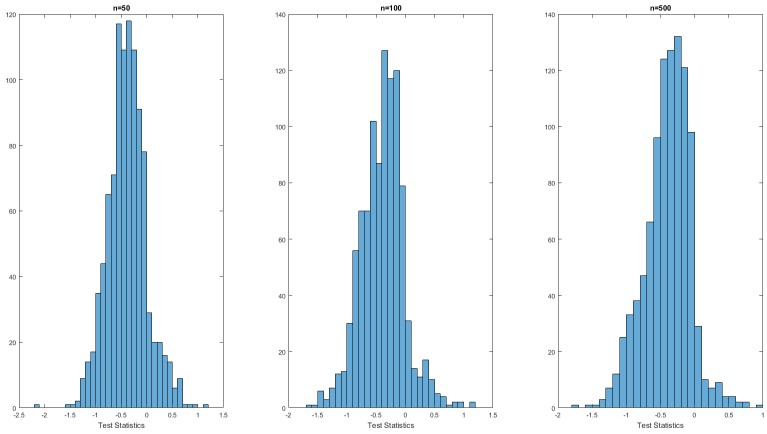

Figure 19: Histograms of V_n under null hypothesis Case I.

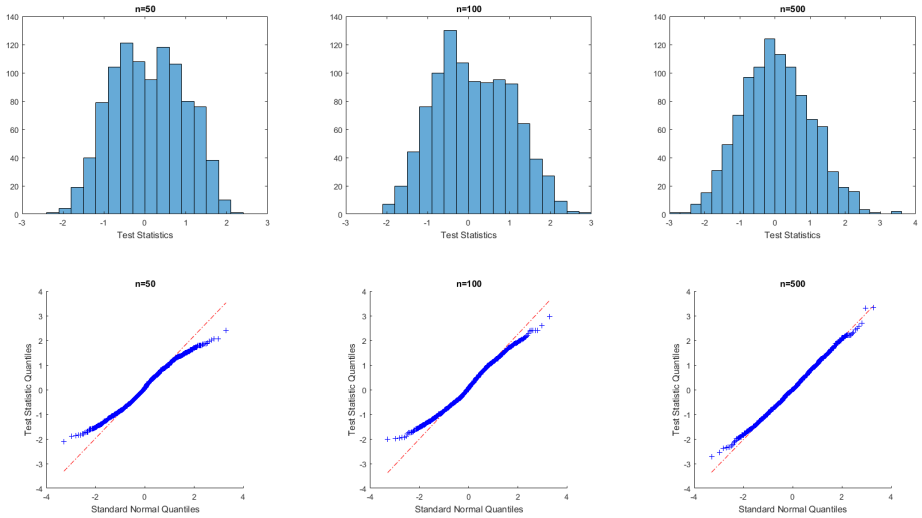

Figure 20: Histograms and quantile-quantile plots to illustrate the convergence of V_n to $N(0,1)$ in distribution under null hypothesis Case II.

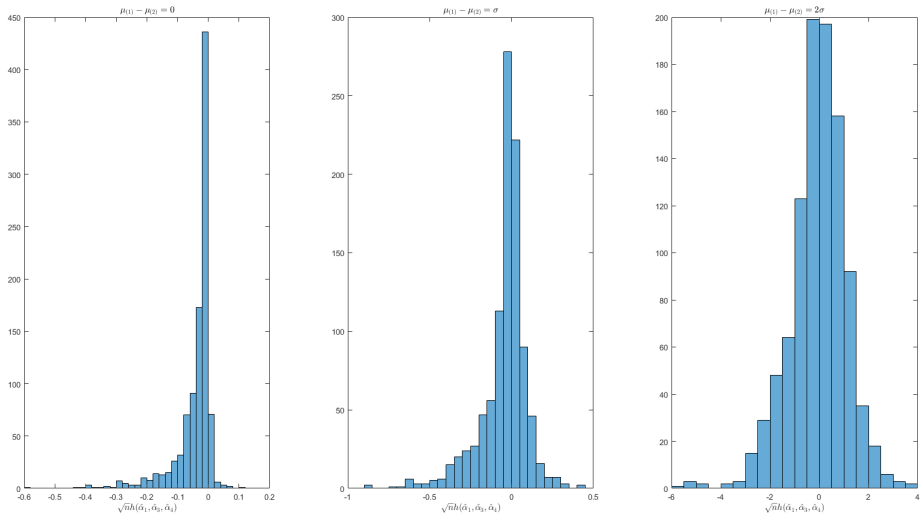

Figure 21: Histograms of 1000 simulated $\sqrt{n}h(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)$ values from samples of size $n = 100$. From left to right, the differences between the two unique values of $\boldsymbol{\mu}$ are $0, \sigma = \sqrt{3/4}$, and $2\sigma = \sqrt{3}$.

Table 3: Empirical Type I Error Rates Case I

Asym. α	$n = 50$	$n = 100$	$n = 500$
0.01	0	0	0
0.05	0	0	0
0.10	0	0	0

Table 4: Empirical Type I Error Rates Case II

Asym. α	$n = 50$	$n = 100$	$n = 500$
0.01	0.0010	0.0010	0.0070
0.05	0.0270	0.0560	0.0560
0.10	0.1040	0.1150	0.1070

When $\mu = (\mu_1, \mu_2, \mu_3)^T$ has two unique values, it was shown in the previous subsection that a_0 equals 1 or -1. When $a_0 = 1$, there are three solutions for θ: $\pi/6, 5\pi/6,$ and $3\pi/2$. Switching between these three solutions permutes the elements of $\mu = (\mu_1, \mu_2, \mu_3)^T$ since

$$\mu_1 = \mu_3 = \alpha_1 - \sqrt{\alpha_1^2 - \alpha_3} \text{ and } \mu_2 = \alpha_1 + 2\sqrt{\alpha_1^2 - \alpha_3} \text{ for } \theta = \frac{\pi}{6}$$
$$\mu_2 = \mu_3 = \alpha_1 - \sqrt{\alpha_1^2 - \alpha_3} \text{ and } \mu_1 = \alpha_1 + 2\sqrt{\alpha_1^2 - \alpha_3} \text{ for } \theta = \frac{5\pi}{6}$$
$$\mu_1 = \mu_2 = \alpha_1 - \sqrt{\alpha_1^2 - \alpha_3} \text{ and } \mu_3 = \alpha_1 + 2\sqrt{\alpha_1^2 - \alpha_3} \text{ for } \theta = \frac{3\pi}{2}.$$

When $a_0 = -1$, the three solutions for θ are $\pi/2, 7\pi/6,$ and $11\pi/6$. Switching between these three solutions also permutes the elements of $\mu = (\mu_1, \mu_2, \mu_3)^T$ since

$$\mu_1 = \mu_2 = \alpha_1 + \sqrt{\alpha_1^2 - \alpha_3} \text{ and } \mu_3 = \alpha_1 - 2\sqrt{\alpha_1^2 - \alpha_3} \text{ for } \theta = \frac{\pi}{2}$$
$$\mu_1 = \mu_3 = \alpha_1 + \sqrt{\alpha_1^2 - \alpha_3} \text{ and } \mu_2 = \alpha_1 - 2\sqrt{\alpha_1^2 - \alpha_3} \text{ for } \theta = \frac{7\pi}{6}$$
$$\mu_2 = \mu_3 = \alpha_1 + \sqrt{\alpha_1^2 - \alpha_3} \text{ and } \mu_1 = \alpha_1 - 2\sqrt{\alpha_1^2 - \alpha_3} \text{ for } \theta = \frac{11\pi}{6}.$$

When $\mu = (\mu_1, \mu_2, \mu_3)^T$ has 3 unique values, $a_0 \in (-1, 1)$. One solution for θ will be the quantity $\theta^* = \sin^{-1}(a_0)/3 \pmod{2\pi}$. Since the image of the arcsine function is $[-\pi/2, \pi/2]$, it follows that θ^* will belong to $[0, \pi/6) \cup (11\pi/6, 2\pi)$. We have excluded the endpoints $\pi/6$ and $11\pi/6$ because we are assuming that μ has 3 unique values.

If $\theta^* \in [0, \pi/6)$, the other solutions to (4.28) are $\pi - \theta^*$, $\theta^* + 2\pi/3$, $\pi/3 - \theta^*$, $\theta^* + 4\pi/3$, and $5\pi/3 - \theta^*$. Let $\mu_1(\theta)$, $\mu_2(\theta)$, and $\mu_3(\theta)$ denote the equations from (4.13), (4.14), and (4.15), respectively. Switching between these six solutions to (4.28) permutes the elements of μ since it can be shown that

$$\mu_1(\pi - \theta^*) = \mu_2(\theta^*), \mu_2(\pi - \theta^*) = \mu_1(\theta^*), \mu_3(\pi - \theta^*) = \mu_3(\theta^*)$$
$$\mu_1(\theta^* + 2\pi/3) = \mu_2(\theta^*), \mu_2(\theta^* + 2\pi/3) = \mu_3(\theta^*), \mu_3(\theta^* + 2\pi/3) = \mu_1(\theta^*)$$
$$\mu_1(\pi/3 - \theta^*) = \mu_3(\theta^*), \mu_2(\pi/3 - \theta^*) = \mu_2(\theta^*), \mu_3(\pi/3 - \theta^*) = \mu_1(\theta^*)$$
$$\mu_1(\theta^* + 4\pi/3) = \mu_3(\theta^*), \mu_2(\theta^* + 4\pi/3) = \mu_1(\theta^*), \mu_3(\theta^* + 4\pi/3) = \mu_2(\theta^*)$$
$$\mu_1(5\pi/3 - \theta^*) = \mu_1(\theta^*), \mu_2(5\pi/3 - \theta^*) = \mu_3(\theta^*), \mu_3(5\pi/3 - \theta^*) = \mu_2(\theta^*).$$

If $\theta^* \in (11\pi/6, 2\pi)$, the other solutions to (4.28) are $3\pi - \theta^*$, $\theta^* - 2\pi/3$, $11\pi/3 - \theta^*$, $\theta^* - 4\pi/3$, and $7\pi/3 - \theta^*$. Switching between these six solutions permutes the elements of μ since it can be shown that

$$\begin{aligned}
&\mu_1(3\pi - \theta^*) = \mu_2(\theta^*), \mu_2(3\pi - \theta^*) = \mu_1(\theta^*), \mu_3(3\pi - \theta^*) = \mu_3(\theta^*) \\
&\mu_1(\theta^* - 2\pi/3) = \mu_3(\theta^*), \mu_2(\theta^* - 2\pi/3) = \mu_1(\theta^*), \mu_3(\theta^* - 2\pi/3) = \mu_2(\theta^*) \\
&\mu_1(11\pi/3 - \theta^*) = \mu_1(\theta^*), \mu_2(11\pi/3 - \theta^*) = \mu_3(\theta^*), \mu_3(11\pi/3 - \theta^*) = \mu_2(\theta^*) \\
&\mu_1(\theta^* - 4\pi/3) = \mu_2(\theta^*), \mu_2(\theta^* - 4\pi/3) = \mu_3(\theta^*), \mu_3(\theta^* - 4\pi/3) = \mu_1(\theta^*) \\
&\mu_1(7\pi/3 - \theta^*) = \mu_3(\theta^*), \mu_2(7\pi/3 - \theta^*) = \mu_2(\theta^*), \mu_3(7\pi/3 - \theta^*) = \mu_1(\theta^*).
\end{aligned}$$

Thus, if μ has at least two distinct entries and is unknown, the entries of μ can only be estimated up to permutations.

We define $\mu^{perm}(\alpha_1, \alpha_3, \theta^*)$, which is a permutation of μ, as

$$\mu^{perm}(\alpha_1, \alpha_3, \theta^*) = \begin{cases} (\alpha_1, \alpha_1, \alpha_1)^T & \text{if } \mu_1 = \mu_2 = \mu_3 \\ (\mu_1(\alpha_1, \alpha_3, \theta^*), \mu_2(\alpha_1, \alpha_3, \theta^*), \mu_3(\alpha_1, \alpha_3, \theta^*))^T & \text{if } \mu_i \neq \mu_j \text{ for } i, j \end{cases},$$

where θ^* is defined as $\sin^{-1}(a_0)/3$ and $\mu_1(\alpha_1, \alpha_3, \theta^*), \mu_2(\alpha_1, \alpha_3, \theta^*)$, and $\mu_3(\alpha_1, \alpha_3, \theta^*)$ are the equations from (4.13), (4.14), and (4.15), respectively, evaluated at α_1, α_3, and θ^*.

Given $Y_1, \ldots, Y_n$, a random sample of vectorized log-eigenvalues from observed eigendecompositions, let $\hat{\alpha}_1, \hat{\alpha}_3$, and $\hat{\alpha}_4$ denote the sample estimators for α_1, α_3, and α_4 from the previous subsection. The "plug-in" estimator $\mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*)$ with entries

$$\begin{aligned}
\mu_1^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*) &= \hat{\alpha}_1 + \sqrt{3(\hat{\alpha}_1^2 - \hat{\alpha}_3)}\left(-\cos(\hat{\theta}^*) + \frac{1}{\sqrt{3}}\sin(\hat{\theta}^*)\right) \\
\mu_2^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*) &= \hat{\alpha}_1 + \sqrt{3(\hat{\alpha}_1^2 - \hat{\alpha}_3)}\left(\cos(\hat{\theta}^*) + \frac{1}{\sqrt{3}}\sin(\hat{\theta}^*)\right) \\
\mu_3^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*) &= \hat{\alpha}_1 - 2\sqrt{(\hat{\alpha}_1^2 - \hat{\alpha}_3)}\sin(\hat{\theta}^*),
\end{aligned}$$

where $\hat{\theta}^*$ satisfies

$$\sin(3\hat{\theta}^*) = \hat{a}_0 = \frac{\hat{\alpha}_4 - \hat{\alpha}_1^3 + 3\hat{\alpha}_1(\hat{\alpha}_1^2 - \hat{\alpha}_3)}{2(\hat{\alpha}_1^2 - \hat{\alpha}_3)^{3/2}},$$

is a starting point for a sample-based estimator of $\mu^{perm}(\alpha_1, \alpha_3, \theta^*)$.

To avoid complex-valued solutions for $\mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)$, we must modify the estimator to handle the cases (i) $\hat{\alpha}_1^2 - \hat{\alpha}_3 < 0$ and (ii) $\hat{\alpha}_1^2 - \hat{\alpha}_3 \geq 0$ and $|\hat{a}_0| > 1$. Under case (i), we will replace $\hat{\alpha}_1^2 - \hat{\alpha}_3$ with 0, which sets $\mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)$ equal to $\hat{\alpha}_1(1,1,1)^T$. Under case (ii), if $\hat{a}_0 < -1$, we will set $\hat{a}_0$ equal to -1 by setting $\hat{\theta}^*$ equal to $-\pi/6$, which yields the estimators

$$\mu_2^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, -\pi/6) = \mu_3^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, -\pi/6) = \hat{\alpha}_1 + \sqrt{\hat{\alpha}_1^2 - \hat{\alpha}_3}$$
$$\mu_1^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, -\pi/6) = \hat{\alpha}_1 - 2\sqrt{\hat{\alpha}_1^2 - \hat{\alpha}_3}.$$

If $\hat{a}_0 > 1$, we will set $\hat{a}_0$ equal to 1 by setting $\hat{\theta}^*$ equal to $\pi/6$, which yields the estimators

$$\mu_1^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \pi/6) = \mu_3^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \pi/6) = \hat{\alpha}_1 - \sqrt{\hat{\alpha}_1^2 - \hat{\alpha}_3}$$
$$\mu_2^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \pi/6) = \hat{\alpha}_1 + 2\sqrt{\hat{\alpha}_1^2 - \hat{\alpha}_3}.$$

To summarize, the modified sample-based estimator for $\mu^{perm}(\alpha_1, \alpha_3, \theta^*)$ is

$$\mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*) = \begin{cases} \hat{\alpha}_1(1,1,1)^T & \text{if } \hat{\alpha}_1^2 - \hat{\alpha}_3 < 0 \\ \mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, -\pi/6) & \text{if } \hat{\alpha}_1^2 - \hat{\alpha}_3 \geq 0 \text{ and } \hat{a}_0 < -1 \\ \mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \pi/6) & \text{if } \hat{\alpha}_1^2 - \hat{\alpha}_3 \geq 0 \text{ and } \hat{a}_0 > 1 \\ \mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*) & \text{if } \hat{\alpha}_1^2 - \hat{\alpha}_3 \geq 0 \text{ and } |\hat{a}_0| \leq 1. \end{cases} \tag{4.29}$$

This estimator will be a consistent estimator of $\mu^{perm}(\alpha_1, \alpha_3, \theta^*)$, regardless of the multiplicity pattern of μ. First, suppose that $\mu_1 = \mu_2 = \mu_3$. Since $\hat{\alpha}_1, \hat{\alpha}_3,$ and $\hat{\alpha}_4$ are consistent estimators of $\alpha_1, \alpha_3,$ and α_4, it follows that $\sqrt{\hat{\alpha}_1^2 - \hat{\alpha}_3} \xrightarrow{p} \sqrt{\alpha_1^2 - \alpha_3} = 0$, which implies that $\mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*) \xrightarrow{p} \mu^{perm}(\alpha_1, \alpha_3, \theta^*)$ for all four cases on the previous page.

Next, suppose that μ has two unique entries and that $a_0 = 1$. Since $\sqrt{\hat{\alpha}_1^2 - \hat{\alpha}_3} \xrightarrow{p} \sqrt{\alpha_1^2 - \alpha_3} > 0$, it follows that the probability of the event "$\hat{\alpha}_1^2 - \hat{\alpha}_3 < 0$" will converge to 0. The consistency of $\hat{\alpha}_1, \hat{\alpha}_3,$ and $\hat{\alpha}_4$ implies that $\hat{a}_0 \xrightarrow{p} a_0 = 1$, which implies that the probability of the event "$\hat{\alpha}_1^2 - \hat{\alpha}_3 \geq 0$ and $\hat{a}_0 < -1$" will converge to 0. From the consistency of $\hat{\alpha}_1$ and $\hat{\alpha}_3$, we have that $\mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \pi/6) \xrightarrow{p} \mu^{perm}(\alpha_1, \alpha_3, \pi/6)$. Finally, $\hat{a}_0 \xrightarrow{p} 1$ implies that $\hat{\theta}^* \xrightarrow{p} \sin^{-1}(1)/3 = \pi/6$, implying that $\mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*) \xrightarrow{p} \mu^{perm}(\alpha_1, \alpha_3, \pi/6)$. Similar reasoning can be used to argue consistency when $a_0 = -1$.

Finally, suppose that μ has 3 unique values. From the consistency of $\hat{\alpha}_1, \hat{\alpha}_3$, and $\hat{\alpha}_4$, we have that $\sqrt{\hat{\alpha}_1^2 - \hat{\alpha}_3} \xrightarrow{p} \sqrt{\alpha_1^2 - \alpha_3} > 0$, $\hat{a}_0 \xrightarrow{p} a_0 \in (-1, 1)$, and $\hat{\theta}^* \xrightarrow{p} \theta^*$, which imply that the probability of the event "$\hat{\alpha}_1^2 - \hat{\alpha}_3 \geq 0$ and $|\hat{a}_0| \leq 1$" will converge to 1, under which $\mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*) \xrightarrow{p} \mu^{perm}(\alpha_1, \alpha_3, \theta^*)$.

To demonstrate the consistency of $\mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*)$, we have performed simulation experiments drawing 1000 samples of size $n = 50, 100,$ and 500 from trivariate normal distributions $N_3(\boldsymbol{\mu}_1, \sigma^2 I_3), N_3(\boldsymbol{\mu}_2, \sigma^2 I_3), N_3(\boldsymbol{\mu}_3, \sigma^2 I_3)$, and $N_3(\boldsymbol{\mu}_4, \sigma^2 I_3)$. Let $\mu_{(1)} \geq \mu_{(2)} \geq \mu_{(3)}$ denote the ordered values of $\boldsymbol{\mu}$. The variance is $\sigma^2 = 3/4$, the mean $\boldsymbol{\mu}_1$ has one unique value $\mu_{(1)} = \mu_{(2)} = \mu_{(3)} = 2$, the mean $\boldsymbol{\mu}_2$ has two unique values $\mu_{(1)} = \mu_{(2)} = 2 + \sqrt{3} > \mu_{(3)} = 2$, the mean $\boldsymbol{\mu}_3$ has two unique values $\mu_{(1)} = 2 > \mu_{(2)} = \mu_{(3)} = 2 - \sqrt{3}$, and the mean $\boldsymbol{\mu}_4$ has three unique values $\mu_{(1)} = 2 + \sqrt{3} > \mu_{(2)} = 2 > \mu_{(3)} = 2 - \sqrt{3}$. To simulate each observation as the log-eigenvalues from a randomly chosen eigen-decomposition, the components of each observation were permuted by a permutation matrix randomly selected from the uniform distribution on the set of 3×3 permutation matrices. For each sample of observed log-eigenvalues, we computed $\mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*)$ using the formula from Equation 4.29, resulting in 1000 estimates for each combination of means $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3$, and $\boldsymbol{\mu}_4$ with the sample sizes $50, 100,$ and 500. For all of the eigenvalue mean multiplicity patterns presented in Table 5, the empirical MSE decreases as n increases.

4.5 DISCUSSION

The hypothesis tests from sections 4.1 and 4.3 can be used sequentially for classification of the shape of the population log-eigenvalue mean of a random sample of diffusion tensors. After specifying a significance level of $\alpha^{(1)}$, if you fail to reject the null hypothesis of an isotropic population log-eigenvalue mean at level $\alpha^{(1)}$, classify the mean as isotropic. If you reject the null hypothesis of an isotropic population log-eigenvalue mean at level $\alpha^{(1)}$, set a second significance level $\alpha^{(2)}$ for a second round of testing. If you fail to reject the null hypothesis that the population log-eigenvalue mean has at most two unique values at level $\alpha^{(2)}$, then classify the population log-eigenvalue mean as oblate if $\hat{a}_0$ is closer to -1 or prolate

if $\hat{a}_0$ is closer to 1. If you reject the null hypothesis that the population log-eigenvalue mean has at most two unique values at level $\alpha^{(2)}$, then classify the population log-eigenvalue mean as triaxial. More work needs to be done to assess the potential classification accuracy of this approach.

While the methods presented in this chapter make no assumptions about the generating distribution of the log-eigenvalues and are based on sample statistics which are easy to compute, the assumption that the generating log-eigenvalue distribution have an isotropic covariance pattern is a strong assumption. More work needs to be done to see if these methods can be extended to allow for more general covariance patterns in the generating log-eigenvalue distribution. A good start for this might be the compound symmetry pattern in which the covariance matrix is parametrized by a single variance term and a single covariance term.

It might also be of interest to perform shape classification of population eigenvalue means over a region of the brain, rather than at a single voxel, from brain scans from n individuals. We would need to adjust the shape classification procedure to account for multiple tests over the region of interest. At each voxel, it would be interesting to approximate the probability that the population mean is isotropic, oblate, prolate, or triaxial, and then color the voxel using a weighted average of four colors. Voxels at which the mean is nearly isotropic, oblate, etc. would be closest in color to the color corresponding to isotropic, oblate, etc.

Table 5: Empirical Mean Squared Error (MSE) from 1000 Simulations of $\mu^{perm}(\hat{\alpha}_1, \hat{\alpha}_3, \hat{\theta}^*)$

$\boldsymbol{\mu}$	$n = 50$	$n = 100$	$n = 500$
$\mu_{(1)} = \mu_{(2)} = \mu_{(3)}$	0.2134	0.1546	0.0652
$\mu_{(1)} = \mu_{(2)} > \mu_{(3)}$	0.1487	0.1001	0.0345
$\mu_{(1)} > \mu_{(2)} = \mu_{(3)}$	0.1464	0.0951	0.0347
$\mu_{(1)} > \mu_{(2)} > \mu_{(3)}$	0.0639	0.0318	0.0062

5.0 PARAMETRIC ESTIMATION AND INFERENCE ON THE SPACE OF EIGEN-DECOMPOSITIONS

In this chapter, we assume that we observe a random sample $S_1, \ldots, S_n \in Sym^+(p)$ which arise via eigen-composition of latent generating eigen-decompositions $(D_1, U_1), \ldots, (D_n, U_n)$. If the unobservable generating eigen-decompositions follow a distribution with location parameter (D^*, U^*), what can we learn about (D^*, U^*) from our random sample of SPD matrices? In this chapter, we present a likelihood-based approach to estimation and inference about the location parameter (D^*, U^*) that allows the user to specify the distribution of the eigenvalues and eigenvectors, and allows for direct estimation and inference of eigenvalue and eigenvector parameters.

5.1 EXAMPLES OF GENERATING DISTRIBUTIONS

We first provide some examples of generating distributions on $Diag^+(p) \times SO(p)$ with location parameter (D^*, U^*) for $p = 2$. Let $\theta_U = sign(U_{21}) \cos^{-1}(U_{11})$ denote the unique angle parametrization of $U \in SO(2)$.

We first present a distribution F_1 in which D and U are independent with $Log(D) \sim N_2(Log(D^*), \tau_1^{-1} I_2)$ and θ_U following the Von Mises distribution with mean θ_{U^*} and concentration parameter $\tau_2 > 0$. The density function of F_1 is given by

$$f_1((D,U)) = \left(\frac{\tau_1}{2\pi}\right) det(D^{-1}) \exp\left\{ -\frac{\tau_1}{2} d^2_{\mathcal{D}^+(p)}(D, D^*) \right\} \frac{1}{2\pi I_0(\tau_2)} \exp\{\tau_2 \cos(\theta_U - \theta_{U^*})\}$$

where $I_0(.)$ denotes the modified Bessel function of order 0. Use of the Von-Mises distribution is prolific in the field of directional statistics, and more information about it can be found in [35].

Alternatively, one could allow θ_U to be the circular version of a normal random variable with mean θ_{U^*} and variance τ_2^{-1} truncated to lie in the interval $(\theta_{U^*} - \pi, \theta_{U^*} + \pi)$ and again have D, which is indepdent of θ_U, follow the log-normal distribution satisfying $Log(D) \sim N_2(Log(D^*), \tau_1^{-1} I_2)$. The density function for this distribution F_2 is

$$f_2((D,U)) = \left(\frac{\tau_1}{2\pi}\right) det(D^{-1}) \exp\left\{ -\frac{\tau_1}{2} d^2_{\mathcal{D}^+(p)}(D, D^*) \right\} \frac{1}{N(\tau_2)} \exp\left\{ -\frac{\tau_2}{2} d^2_C(\theta_U, \theta_{U^*}) \right\}$$

where $N(\tau_2) = \sqrt{\frac{2\pi}{\tau_2}} erf(\pi \sqrt{\frac{\tau_2}{2}})$ and $d_C(.,.)$ denotes the circular minimal arc length distance function given by

$$d_C(\theta_1, \theta_2) = \begin{cases} |\bar{\theta}_1 - \bar{\theta}_2| & \text{if } |\bar{\theta}_1 - \bar{\theta}_2| \leq \pi \\ |\bar{\theta}_1 - \bar{\theta}_2| - \pi & \text{if } |\bar{\theta}_1 - \bar{\theta}_2| > \pi \end{cases}$$

with $\bar{\theta}_i = \theta_i \pmod{2\pi}$ for $i = 1, 2$. We use the circular distance function $d_C(.,.)$ to ensure that the distribution of θ_U is periodic (i.e. invariant under shifts by integer multiples of 2π). It can be shown that

$$d_C(\theta_1, \theta_2) = d_{SO(2)}(U_{\theta_1}, U_{\theta_2})$$

where $U_{\theta_i} \in SO(2)$ is the 2×2 rotation matrix associated with angle θ_i and $d_{SO(2)}$ is the Riemannian affine-invariant distance function from Equation 2.8. Many properties of this circular truncated normal distribution are presented in [12].

Finally, we present a distribution F_3 that allows for dependence between the scaling and rotation variables. Let $v((D,U);(D^*,U^*)) = vec(Log(D(D^*)^{-1}), Log(U(U^*)^T))$, a 3-dimensional vector containing the free entries from $(Log(D(D^*)^{-1}), Log(U(U^*)^T))$, which are the coordinates of (D,U) on the tangent space centered at (D^*,U^*) (see Theorem 2.8.1). To compute $v((D,U);(D^*,U^*))$, note that

$$Log(D(D^*)^{-1}) = \begin{pmatrix} \log(d_1) - \log(d_1^*) & 0 \\ 0 & \log(d_2) - \log(d_2^*) \end{pmatrix}$$

$$Log(U(U^*)^T) = \begin{pmatrix} 0 & -\theta(U(U^*)^T) \\ \theta(U(U^*)^T) & 0 \end{pmatrix}$$

where $\theta(U(U^*)^T) = sgn([U(U^*)^T]_{21}) \cos^{-1}([U(U^*)^T]_{11})$. Then we have that

$$v((D,U);(D^*,U^*)) = (\log(d_1) - \log(d_1^*), \log(d_2) - \log(d_2^*), \theta(U(U^*)^T))^T.$$

We then define F_3 as a kind of tangent normal distribution with density function

$$f_3((D,U)) = \frac{1}{N(\Sigma)} \exp\left\{ -\frac{1}{2}(v((D,U);(D^*,U^*)))^T \Sigma^{-1} v((D,U);(D^*,U^*)) \right\}$$

where $N(\Sigma)$ is the normalizing constant and $\Sigma \in Sym^+(3)$ defines the covariance structure on the tangent space. For highly concentrated distributions of this type, we make the simplification

$$N(\Sigma) \approx (det(2\pi\Sigma))^{1/2}$$

from approximating the distribution of $v((D,U);(D^*,U^*))$ with a mean zero trivariate normal distribution with covariance matrix Σ.

All three of these distributions on $Diag^+(2) \times SO(2)$ can be generalized for any $p \geq 2$.

5.2 MIXTURE-BASED LIKELIHOOD ESTIMATION FRAMEWORK

Given a random sample $S_1, \ldots, S_n$ of SPD matrices, we assume that the observations arise as eigen-compositions of unobservable i.i.d. generating eigen-decompositions. Our goal is to estimate the parameters of the generating eigen-decomposition distribution from a sample of SPD matrices. We make the following assumptions:

1. The unobservable generating eigen-decompositions $(D_1, U_1), \ldots, (D_n, U_n)$ are i.i.d. observations from a continuous distribution F on $Diag^+(p) \times SO(p)$ with location parameter (D^*, U^*) and concentration parameters Γ. F has density function $f((D,U);(D^*,U^*),\Gamma)$.

2. The generating distribution F is a location family: if a random variable (D, U) follows distribution F with location parameter (D^*, U^*) and concentration parameters Γ, then the random variable $(D', U') = (GDG^T, UG^T)$ follows distribution F with location parameter (GD^*G^T, U^*G^T) and concentration parameters Γ for any even signed permutation matrix $G \in \mathcal{G}(p)$.
3. Observed eigen-decompositions $(D'_1, U'_1), \ldots, (D'_n, U'_n)$ of $S_1, \ldots, S_n$ are randomly chosen from the uniform distributions on $\mathcal{F}^{-1}(S_1), \ldots, \mathcal{F}^{-1}(S_n)$.

From these assumptions, we have that each observation will have no repeated eigenvalues with probability 1, and that the distribution of an observed eigen-decomposition will be a mixture distribution with density function

$$h((D', U'); (D^*, U^*), \Gamma) = \frac{1}{2^{p-1}p!} \sum_{j=1}^{2^{p-1}p!} f((D', U'); (G_j D^* G_j^T, U^* G_j^T), \Gamma). \tag{5.1}$$

Note that the location parameter (D^*, U^*) is not identifiable since

$$h((D', U'); (D^*, U^*), \Gamma) = h((D', U'); (GD^*G^T, U^*G^T), \Gamma)$$

for any $G \in \mathcal{G}(p)$. Thus, the location parameter of the unobservable generating distribution is identifiable up to signed permutations; this is equivalent to defining our parameter space of interest as the quotient group $(Diag^+(p) \times SO(p))/\mathcal{G}(p)$.

As a simple illustration, suppose that $(D_1, U_1), \ldots, (D_n, U_n)$ is an i.i.d. random sample on $Diag^+(2) \times SO(2)$ (plotted in black in the left plot from Figure 22), which yields a random sample $\mathcal{F}(D_1, U_1) = S_1, \ldots, \mathcal{F}(D_n, U_n) = S_n$ of 2×2 SPD matrices (plotted in light blue in the right plot from Figure 22). The three other versions of the eigen-decompositions of $S_1, \ldots, S_n$ are plotted in red, blue, and green in the left plot from Figure 22. When extracting the "observed" eigen-decompositions $(D'_1, U'_1), \ldots, (D'_n, U'_n)$ from our sample of SPD matrices $S_1, \ldots, S_n$, we will never know exactly how the observed eigen-decomposition (D'_i, U'_i) is related to the unobserved data generating eigen-decomposition (D_i, U_i) for $i = 1, \ldots, n$. In relation to Figure 22, we will never know if eigen-decomposition (D'_i, U'_i) corresponds to a point from the black, red, blue, or green point cloud from the left plot of all possible

eigen-decompositions. Under the mixture model from Equation 5.1, an observed eigen-decomposition occurs as a uniform random draw from the black, red, blue, or green point cloud. To make sure that observed eigen-decompositions follow this mixture model in practice, we will select an intial eigen-decomposition for each observation S_i and then map that initial eigen-decomposition to (possibly) another eigen-decomposition of S_i via a randomly selected even signed permutation from the uniform distribution on $\mathcal{G}(p)$.

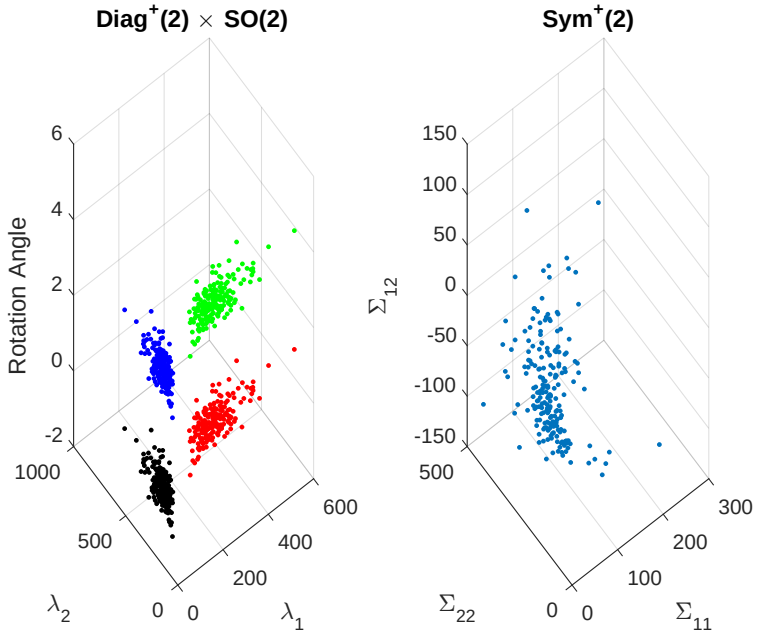

Figure 22: Plots of a random sample of 2x2 SPD matrices (right) and their 4 different eigen-decompositions in the cross-product space (left). The different versions of the eigen-decompostions are represented by the four distinct colors.

We plan to perform parameter estimation and inference in this setting via likelihood-based methods. From Equation 5.1, the log-likelihood of the observed eigen-decompositions has the form

$$\ell((D^*, U^*), \Gamma) = \sum_{i=1}^{n} \log \left\{ \frac{1}{2^{p-1}p!} \sum_{j=1}^{2^{p-1}p!} f((G_j^T D_i' G_j, U_i' G_j); (D^*, U^*), \Gamma) \right\}. \tag{5.2}$$

Optimizing Equation 5.2 directly may be computationally intractable, however the mixture distribution of the observed eigen-decompositions lends suggests that the EM algorithm may be useful for indirectly maximizing the likelihood above.

5.3 MEAN ESTIMATION VIA EM ALGORITHM

For each observation S_i, label its eigen-decompositions as

$$\mathcal{F}^{-1}(S_i) = \bigcup_{j=1}^{2^{(p-1)}p!} \{(D_i^{(j)}, U_i^{(j)})\},$$

where versions $(D_1^{(j)}, U_1^{(j)}), \ldots, (D_n^{(j)}, U_n^{(j)})$ form an i.i.d. sample from a continuous distribution F with location parameter $(\mathcal{D}^{(j)}, \mathcal{U}^{(j)})$ satisfying assumptions 1-3 from the previous section. To simplify the notation of the set of location parameters, let $(\mathcal{D}^{(1)}, \mathcal{U}^{(1)}) = (D^*, U^*)$. By assumption 2 from the previous section, $(\mathcal{D}^{(j)}, \mathcal{U}^{(j)}) = (G_j D^* G_j^T, U^* G_j^T)$ for some $G_j \in \mathcal{G}(p)$ for $j = 1, \ldots, 2^{(p-1)}p!$. Then we will denote the set of location parameters as $[(D^*, U^*)] \in (Diag^+(p) \times SO(p))/\mathcal{G}(p)$, which is the orbit of (D^*, U^*) under the group action of $\mathcal{G}(p)$ on $Diag^+(p) \times SO(p)$ (as defined in subsection 3.2.1).

Let $k(p) = 2^{(p-1)}p!$, and suppose that $(D_1', U_1'), \ldots, (D_n', U_n')$ is a sample of observed eigen-decompositions as specified in assumption 3 from the previous section. Each observed eigen-decomposition has the mixture representation

$$(D_i', U_i') = \sum_{j=1}^{k(p)} (Z_j D_i^{(j)}, Z_j U_i^{(j)})$$

where $\bar{Z} = (Z_1, \ldots, Z_{k(p)})$ follows a multinomial distribution with the constraint $\sum_{j=1}^{k(p)} Z_i = 1$ and probabilities $\bar{\pi} = (1/k(p), \ldots, 1/k(p))$. This representation leads to the observed data log-likelihood from Equation 5.2, which will be difficult to optimize in practice.

If we were able to observe the version indicator variable z_{ij} for each observed eigen-decomposition (D_i', U_i'), then we could form the complete data log-likelihood

$$\ell_C((D^*, U^*), \Gamma) = \sum_{i=1}^{n} \sum_{j=1}^{k(p)} z_{ij} \log[f((G_j^T D_i' G_j, U_i' G_j); (D^*, U^*), \Gamma)] - n \log(k(p))$$

and estimate the parameters using standard maximum likelihood estimation. The EM algorithm will address our missing data problem by iteratively imputing values for the unobservable version indicator variables and then performing maximum likelihood estimation using the complete data log-likelihood with imputed values plugged in for the version indicator variables. The steps are:

1. Choose initial estimates $(\hat{D}^{*(0)}, \hat{U}^{*(0)})$ and $\hat{\Gamma}^{(0)}$ for the parameters (D^*, U^*) and Γ and set tolerance $\varepsilon > 0$.
2. Expectation Step: Compute

$$\hat{z}_{ij}^{(k)} = \frac{f((G_j^T D_i' G_j, U_i' G_j); (\hat{D}^{*(k)}, \hat{U}^{*(k)}), \hat{\Gamma}^{(k)})}{\sum_{j=1}^{k(p)} f((G_j^T D_i' G_j, U_i' G_j); (\hat{D}^{*(k)}, \hat{U}^{*(k)}), \hat{\Gamma}^{(k)})}$$

3. Maximization Step: Compute

$$((\hat{D}^{*(k+1)}, \hat{U}^{*(k+1)}), \hat{\Gamma}^{(k+1)}) = \text{argmax} \sum_{i=1}^{n} \sum_{j=1}^{k(p)} \hat{z}_{ij}^{(k)} \log[f((G_j^T D_i' G_j, U_i' G_j); (D^*, U^*), \Gamma)]$$

4. If $|\ell((\hat{D}^{*(k+1)}, \hat{U}^{*(k+1)}), \hat{\Gamma}^{(k+1)}) - \ell((\hat{D}^{*(k)}, \hat{U}^{*(k)}), \hat{\Gamma}^{(k)})| > \varepsilon$, then return to Step 2. Otherwise, set the final parameter estimates equal to $\hat{\Phi}^{(k+1)}$.

We illustrate how to perform parameter estimation via the EM algorithm when the generating eigen-decompositions follow distribution F_2 with $p = 2$ from Section 5.1 with density function

$$f_2((D, U); (D^*, U^*), \Gamma) = \left(\frac{\gamma_1}{2\pi N(\gamma_2)|D|} \right) \exp \left\{ -\frac{\gamma_1}{2} d^2_{\mathcal{D}^+(2)}(D, D^*) - \frac{\gamma_2}{2} d^2_{SO(2)}(U, U^*) \right\}$$

where $|D| = det(D)$ and $N(\gamma_2) = \sqrt{2\pi/\gamma_2} erf(\pi\sqrt{\gamma_2/2})$. Let $S_1, \ldots, S_n$ be a random sample of SPD matrices arising from eigen-composition of the distribution F_2 and let (D_i', U_i') denote an observed eigen-decomposition of S_i randomly drawn from the uniform distribution on $\mathcal{F}^{-1}(S_i)$. We begin with initial guesses $(\hat{D}^{*(0)}, \hat{U}^{*(0)})$ and $\hat{\Gamma}^{(0)}$ for the parameters (D^*, U^*) and Γ.

Performing the expectation step after k iterations simply requires computing

$$\hat{z}_{ij}^{(k)} = \frac{f_2((G_j^T D_i' G_j, U_i' G_j); (\hat{D}^{*(k)}, \hat{U}^{*(k)}), \hat{\Gamma}^{(k)})}{\sum_{j=1}^{4} f_2((G_j^T D_i' G_j, U_i' G_j); (\hat{D}^{*(k)}, \hat{U}^{*(k)}), \hat{\Gamma}^{(k)})}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, 4$.

Performing the maximization for the $k+1$ iteration requires solving

$$\hat{D}^{*(k+1)} = \underset{D \in Diag^+(2)}{\text{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{4} \left(\frac{\hat{z}_{ij}^{(k)}}{n} \right) d^2_{\mathcal{D}^+(2)}(G_j^T D_i' G_j, D) \tag{5.3}$$

$$\hat{U}^{*(k+1)} = \underset{U \in SO(2)}{\text{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{4} \left(\frac{\hat{z}_{ij}^{(k)}}{n} \right) d^2_{SO(2)}(U_i' G_j, U) \tag{5.4}$$

to obtain estimates for the location parameters. It can be shown that

$$\hat{D}^{*(k+1)} = \sum_{i=1}^{n}\sum_{j=1}^{4}\left(\frac{\hat{z}_{ij}^{(k)}}{n}\right) G_j^T D_i' G_j$$

and that solving for $\hat{U}^{*(k+1)}$ requires locating $U \in SO(2)$ that satisfy

$$\sum_{i=1}^{n}\sum_{j=1}^{4}\left(\frac{\hat{z}_{ij}^{(k)}}{n}\right) Log(U_i' G_j U^T) = \mathbf{0}.$$

After substituting $(\hat{D}^{*(k+1)}, \hat{U}^{*(k+1)})$ for (D^*, U^*) into the approximate complete data log-likelihood, we can solve the equations

$$\hat{\gamma}_1^{(k+1)} = \underset{\gamma_1}{\text{argmax}}\, n\log(\gamma_1) - \frac{\gamma_1}{2}\sum_{i=1}^{n}\sum_{j=1}^{4}\left(\frac{\hat{z}_{ij}^{(k)}}{n}\right) d^2_{\mathcal{D}^+(2)}(G_j^T D_i' G_j, \hat{D}^{*(k+1)}) \tag{5.5}$$

$$\hat{\gamma}_2^{(k+1)} = \underset{\gamma_2}{\text{argmax}}\, -n\log(N(\gamma_2)) - \frac{\gamma_2}{2}\sum_{i=1}^{n}\sum_{j=1}^{4}\left(\frac{\hat{z}_{ij}^{(k)}}{n}\right) d^2_{SO(2)}(U_i' G_j, \hat{U}^{*(k+1)}) \tag{5.6}$$

to obtain estimates for the concentration parameters. After differentiating Equation 5.5 and setting the derivative equal to 0, we have that

$$\hat{\gamma}_1^{(k+1)} = \left[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{4}\left(\frac{\hat{z}_{ij}^{(k)}}{n}\right) d^2_{\mathcal{D}^+(2)}(G_j^T D_i' G_j, \hat{D}^{*(k+1)})\right]^{-1}.$$

From differentiating Equation 5.6 and setting the derivative equal to 0, we get that $\hat{\gamma}_2^{(k+1)}$ should satisfy

$$V(\hat{\gamma}_2^{(k+1)}) = -\frac{N'(\hat{\gamma}_2^{(k+1)})}{N(\hat{\gamma}_2^{(k+1)})} = \sum_{i=1}^{n}\sum_{j=1}^{4}\left(\frac{\hat{z}_{ij}^{(k)}}{n}\right) d^2_{SO(2)}(U_i' G_j, \hat{U}^{*(k+1)}).$$

It is shown in [12] that V is a one-to-one function, so $\hat{\gamma}_2^{(k+1)}$ is unique and can be approximated with a numerical procedure such as Newton's method.

To demonstrate the performance of this estimation method, we performed simulation experiments drawing 500 samples of size $n = 50, 100,$ and 500 from the distribution F_2 with $p = 2$ and parameters $\gamma_1 = 2$, $\gamma_2 = 5$,

$$D^* = \begin{pmatrix} 10 & 0 \\ 0 & 2 \end{pmatrix}, \text{and } U^* = \begin{pmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{pmatrix}.$$

Table 6: Empirical Mean Squared Error (MSE) from 500 Simulations

Parameter	$n = 50$	$n = 100$	$n = 500$
(D^*, U^*)	1.6042	0.8928	0.1756
γ_1	0.1068	0.0493	0.0089
γ_2	1.7845	0.8069	0.1237

We computed the empirical mean squared error from estimating (D^*, U^*) using the formula

$$MSE((D^*, U^*)) = \frac{1}{500} \sum_{t=1}^{500} \|\mathcal{F}(\hat{D}_t^*, \hat{U}_t^*) - \mathcal{F}(D^*, U^*)\|_F^2$$

where $(\hat{D}_t^*, \hat{U}_t^*)$ is the sample estimate for (D^*, U^*) from sample t. We can see in Table 6 that the empirical MSE for each parameter decreases as n increases, suggesting that the EM estimation procedure is consistent for this example.

5.4 DISCUSSION

It would be interesting to establish conditions on the distribution of the generating eigen-decompositions which would guarantee that the EM algorithm would find a global maximizer of the likelihood function and would provide consistent estimates. Perhaps more interesting, the mixture likelihood-based estimation framework proposed in this chapter can easily be extended to create likelihood ratio tests for single and multi-sample inference. For example, to test the null hypothesis that $(D^*, U^*) = (D_{(0)}, U_{(0)}), \Gamma = \Gamma_0$, one could use the test statistic

$$\Lambda = 2[\ell((\hat{D}^*, \hat{U}^*), \hat{\Gamma}) - \ell((D_0, U_0), \Gamma_0)],$$

where $\ell((\hat{D}^*, \hat{U}^*), \hat{\Gamma})$ is the observed log-likelihood evaluated at the EM parameter estimates and $\ell((D_0, U_0), \Gamma_0)$ is the observed log-likelihood evaluated at the null hypothesis parameter values. Under the null hypothesis, we conjecture that Λ will asymptotically follow a chi-square distribution with degrees of freedom equal to the number of free parameters in $D^*, U^*,$ and Γ.

Given data from two independent samples, it might be of interest to test the null hypothesis that the two samples have the same location parameters (i.e. $(D_1^*, U_1^*) = (D_2^*, U_2^*)$) while assuming that they have the same concentration parameters. For this test, we could use the test statistic

$$\Lambda = 2[\ell((\hat{D}_1^*, \hat{U}_1^*), \hat{\Gamma}_{pooled}) + \ell((\hat{D}_2^*, \hat{U}_2^*), \hat{\Gamma}_{pooled}) - \ell((\tilde{D}^*, \tilde{U}^*), \tilde{\Gamma})],$$

where $(\hat{D}_1^*, \hat{U}_1^*)$ and $(\hat{D}_2^*, \hat{U}_2^*)$ are the EM location estimates for samples 1 and 2 under the alternative hypothesis, $\hat{\Gamma}_{pooled}$ contains the pooled concentration parameter estimates under the alternative hypothesis, and $(\tilde{D}^*, \tilde{U}^*)$ and $\tilde{\Gamma}$ are the EM estimates for the location and concentration parameters under the null hypothesis after combining both samples into one sample. We conjecture that, under the null hypothesis, this test statistic will asymptotically follow a chi-square distribution with degrees of freedom equal to the number of free parameters in D^* and U^*.

BIBLIOGRAPHY

[1] D. Alexander, *Muliple-fibre Reconstruction Algorithms for Diffusion MRI*, Annals of the New York Academy of Sciences 1046 (2005), pp. 113-133.

[2] 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., 20 (2007), pp. 328-347.

[3] B. Asfari, *Riemannian L^p Centers of Mass: Existence, Uniqueness, and Convexity*, Proc. Amer. Math. Soc., 139 (2011), pp. 655-673.

[4] P.G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, A. Connelly, *A Rigorous Framework for Diffusion Tensor Calculus*, Magnet. Reson. Med., 53 (2005), pp. 221-225.

[5] R. N. Bhattacharya and L. Lin, *Omnibus CLTs for Fréchet Means and Non-Parametric Inference on Non-Euclidean Spaces*, Proc. Amer. Math. Soc., 145 (2017), 413-428.

[6] R. N. Bhattacharya and V. Patrangenaru, *Large Sample Theory of Intrinsic and Extrinsic Sample Means on Manifolds-I*, Annals of Statistics, 31 (2003), pp. 1-29.

[7] R. N. Bhattacharya and V. Patrangenaru, *Large Sample Theory of Intrinsic and Extrinsic Sample Means on Manifolds-II*, Annals of Statistics, 33 (2005), pp. 1225-1259.

[8] P. Bickel, E. Levina, *Regularized Estimation of Large Covariance Matrices*, Annals of Statistics, 36 (2008), pp. 199-227.

[9] O. Carmichael, J. Chen, D. Paul, and J. Peng, *Diffusion Tensor Smoothing Through Weighted Karcher Means*, Electronic Journal of Statistics, 7 (2013), pp. 1913-1956.

[10] T.C. Chao, M.C. Chou, P. Yang, H.W. Chung, M.T. Wu, *Effects of Interpolation Methods in Spatial Normalization of Diffusion Tensor Imaging Data on Group Comparisons of Fractional Anisotropy*, Magn. Reson. Imaging, 27 (2009), pp. 681-690.

[11] Y. Chikuse, *Statistics on Special Manifolds*, Springer, New York, 2003.

[12] J-F Coeurjelly, N. Le Bihan, *Geodesic Normal Distribution on the Circle*, Metrika, 75 (2012), pp. 977-995.

[13] M. Daniels, R. Kass, *Shrinkage Estimators for Covariance Matrices*, Biometrics, 57 (2001), pp. 1173-1184.

[14] M. Daniels, M. Pourahmadi, *Bayesian Analysis of Covariance Matrices and Dynamic Models for Longitudinal Data*, Biometrika, 89 (2002), pp. 553-566.

[15] I.L. Dryden, A.A. Kolydenko, and D. Zhou, *Non-Euclidean Statistics for covariance matrices, with applications to diffusion tensor imaging*, Ann. Applied Statistics 3 (2009), pp. 1102-1123.

[16] M. DoCarmo, *Riemannian Geometry*, Birkhauser, Boston, 1992.

[17] P. T. Fletcher, *Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds*, Int J Comput Vis, 105 (2013), pp. 171-185.

[18] P. T. Fletcher, T. Lu, C. Lu, S. Pizer, and S. Joshi, *Principal Geodesic Analysis for the Study of Non-Linear Statistics of Shape*, IEEE Trans on Med Imaging 23 (2004), pp. 995-1005.

[19] M. Fréchet, *Les éléments aléatoires de nature quelconque dans un espace distancié*, Ann. Inst. H. Poincaré Probab. Statist. 10 (1945), pp. 215-310.

[20] J. H. Gallier, *Geometric Methods and Applications: For Computer Science and Engineering*, Texts Appl. Math. 38, Springer, New York, 2011.

[21] J.C. Gower, *Generalized Procrustes Analysis*, Psychometrika, 40 (1975), pp. 33-51.

[22] D. Groisser, S. Jung, A. Schwartzman, *Eigenvalue Stratification and Minimal Smooth Scaling-Rotation Curves in the Space of Symmetric Positive Definitie Matrices*, arXiv:1602.01187v3, 2016.

[23] D. Groisser, S. Jung, A. Schwartzman, *Geometric Foundations for Scaling-Rotation Statistics on Symmetric Positive Definite Matrices: Minimal Smooth Scaling-Rotation Curves in Low Dimensions*, Electronic Journal of Statistics, 11 (2017), pp. 1092-1159.

[24] K. Grove, H. Karcher, E. A. Ruh, *Jacobi Fields and Finsler Metrics on Compact Lie Groups with an Application to Differentiable Pinching Problems*, Mathematische Annalen, 211 (1974), pp. 7-21.

[25] N. Higham, A. Al-Mohy, *Computing Matrix Functions*, Acta Numerica, 19 (2010), pp. 159-208.

[26] S. Huckemann, *Inference on 3D Procrustes Means: Tree Bole Growth, Rank-Deficient Diffusion Tensors, and Perturbation Models*, Scandinavian Journal of Statistics, 38 (2011), pp. 424-446.

[27] S. Huckemann, *Intrinsic Inference on the Mean Geodesic of Planar Shapes and Tree Discrimination by Leaf Growth*, Annals of Statistics, 39 (2011), pp. 1098-1124.

[28] S. Jung, A. Schwartzman, D. Groisser, *Scaling-Rotation Distance and Interpolation of Symmetric Positive-Definite Matrices*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1180-1201.

[29] H. Karcher, *Riemannian Center of Mass and Mollifier Smoothing*, Comm. Pure Appl. Math., 30 (1977), pp. 509-541.

[30] C. Khatri and K. Mardia, *The Von Mises-Fisher Matrix Distribution in Orientation Statistics*, Journal of the Royal Statistical Society Series B, 39 (1977), pp. 95-106.

[31] P.B. Kingsley, *Introduction to Diffusion Tensor Imaging Mathematics: Part II. Anisotropy, Diffusion-weighting Factors, and Gradient Encoding Schemes*, Concepts in Magnetic Resonance Part A 28A (2006), pp. 123-154.

[32] H. Le, *Locating Fréchet Means with Application to Shape Spaces*, Adv. in Appl. Probab., 33 (2001), pp. 324-338.

[33] N. Lepore, C. Brun, M. Chiang, Y. Chou, R. Dutton, K. Hayashi, E. Luders, O. Lopez, H. Aizenstein, A. Toga, J. Becker, and P. Thompson, *Generalized Tensor-Based Morphometry of HIV/AIDS Using Multivariate Statistics on Deformation Tensors*, IEEE Trans Med Imaging, 27 (2008), pp. 129-141.

[34] J. Manton, *A Globally Convergent Numerical Algorithm for Computing the Centre of Mass on Compact Lie Groups*, ICARV 2004 8th Control, Automation, Robotics, and Vision Conference.

[35] K. Mardia and P. Jupp, *Directional Statistics*, John Wiley and Sons Ltd., 2000.

[36] M. Moakher, *Means and Averaging in the Group of Rotations*, SIAM J. Matrix Anal. Appl., 24 (2001), pp. 1-16.

[37] 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.

[38] N. Paquette, J. Shi, Y. Wang, Y. Lao, R. Ceschin, M. D. Nelson, A. Panigrahy, and N. Lepore, *Ventricular Shape and Relative Position Abnormalities in Preterm Neonates*, Neuroimage Clin., 15 (2017), pp. 483-493.

[39] X. Pennec, *Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements*, J. Math Imaging Vis., 25 (2006), pp. 127-154.

[40] X. Pennec, P. Fillard, and N. Ayache, *A Riemannian Framework for Tensor Computing*, Int. J. Comput. Vis., 66 (2006), pp. 41-66.

[41] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 1976.

[42] A. Schwartzman, *Random Ellipsoids and False Discovery Rates: Statistics for Diffusion Tensor Imaging Data*, Ph. D. Thesis, Stanford University, Stanford, California, 2006.

[43] A. Schwartzman, *Lognormal Distribution and Geometric Averages of Symmetric Positive Definite Matrices*, International Statistical Review, 0 (2015), pp. 1-31.

[44] A. Schwartzman, W. Mascarenhas, and J. Taylor, *Inference for Eigenvalues and Eigenvectors of Gaussian Symmetric Matrices*, Annals of Statistics, 36 (2008), pp. 2886-2919.

[45] A. Schwartzman, R. Dougherty, and J. Taylor, *Group Comparisons of Eigenvalues and Eigenvectors of Diffusion Tensors*, Journal of the American Statistical Association, 105 (2010), pp. 588-599.

[46] P. Sundgren, Q. Dong, D. Gomez-Hassan, S. Mukherji, P. Maly, and R. Welsh *Diffusion Tensor Imaging of the Brain: Review of Clinical Applications*, Neuroradiology 46 (2004), pp.273-282.

[47] C. Tench, P. Morgan, M. Wilson, and L. Blumhardt, *White Matter Mapping Using Diffusion Tensor MRI*, Magnetic Resonance in Medicine, 47 (2002), pp. 967-972.

[48] D. Zhou, I.L. Dryden, A.A. Koloydenko, K. Audenaert, L. Bai, *Regularisation, Interpolation, and Visualisation of Diffusion Tension Images Using Non-Euclidean Statistics*, Journal of Applied Statistics, Vol. 43, Issue 5, 2016.

[49] H. Zhu, Y. Li, I. Ibraham, X. Shi, H. An, Y. Chen, W. Gao, W. Lin, D. Rowe, and B. Peterson, *Regression Models for Identifying Noise Sources in Magnetic Resonance Images*, Journal of the American Statistical Association, 104 (2009), pp. 623-637.

[50] H. Ziezold, *On Expected Figures and a Strong Law of Large Numbers for Random Elements in Quasi-Metric Spaces*, Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A, pp. 591-602.