DIFFERENTIAL GEOMETRY FOR MODEL INDEPENDENT ANALYSIS OF IMAGES AND OTHER NON-EUCLIDEAN DATA: RECENT DEVELOPMENTS

RABI BHATTACHARYA AND LIZHEN LIN

In celebration of Chuck’s 70th Birthday

ABSTRACT. This article provides an exposition of recent methodologies for nonparametric analysis of digital observations on images and other non-Euclidean objects. Fréchet means of distributions on metric spaces, such as manifolds and stratified spaces, have played an important role in this endeavor. Apart from theoretical issues of uniqueness of the Fréchet minimizer and the asymptotic distribution of the sample Fréchet mean under uniqueness, applications to image analysis are highlighted. In addition, nonparametric Bayes theory is brought to bear on the problems of density estimation and classification on manifolds.

1. INTRODUCTION

Historically, directional statistics, that is, statistics on spheres, especially $S^2$, has been around for a long time, and there is a great deal of literature on it (See the books by Watson (1983), Mardia and Jupp (2000), Fisher et al. (1987)). Much of that was inspired by a seminal paper by Fisher (1953) proving beyond any reasonable doubt that the earth’s magnetic poles had shifted over geological times. Indeed, the two sets of data that he analyzed, one from the Quaternary period and the other from recent times (1947-48), showed an almost reversal of the directions of the magnetic poles. In addition to this first scientific demonstration of a phenomenon conjectured by some paleontologists, such studies of magnetic poles in fossilized remanent magnetism had an enormous impact on tectonics, essentially validating the theory of continental drift (Irving (1964), Fisher et al. (1987)). There are other important applications of directional statistics, such as designing of windmills based on wind directions, etc. Fisher’s example is presented in Section 9, in comparison with the nonparametric method highlighted in this article.

The advancement of imaging technology and increase in computing prowess have opened up a whole new vista of applications. Medical imaging, for example, is now an essential component of medical practice. Not only have MRIs (magnetic resonance imaging) become routine for diagnosing a plethora of diseases, there are more advanced techniques such as the DTI (diffusion tensor imaging) which measures diffusion coefficients of water molecules in tiny voxels along nerve fibers in the cortex of the brain in order to understand or monitor diseases such as Parkinson’s and Alzheimer’s (Goodlett et al., 2006; Kindlmann et al., 2007; Morra et al., 2000). Beyond medicine, there are numerous applications to morphometrics (Bookstein, 1991), graphics, robotics, and machine vision (Aggarwal et al., 2004; Ma et al., 2005; Veeraraghavan et al., 2005).

Images are geometric objects and their precise mathematical descriptions and identifications in different fields of applications are facilitated by the use of differential geometry. Kendall (1984) and Bookstein (1991) were two pioneers in the geometric description and
statistical analysis of images represented by landmarks on two or three dimensional objects. The spaces of such images, or shapes, are differential manifolds, or \textit{stratified spaces} obtained by gluing together manifolds of different dimensions. In the following sections these spaces are described in detail. Much of the earlier statistical analysis on differential manifolds were parametric in nature, where a distribution $Q$ on a manifold $M$ is assumed to belong to a finite dimensional parametric family; that is, $Q$ is assumed to have a density (with respect a standard distribution, e.g., the volume measure on $M$) which is specified except for the value of a finite dimensional parameter $\theta$ lying in an open subset $\Theta$ of an Euclidean space. The statistician’s task is then to estimate the parameter (or test for its belonging to a particular subset of $\Theta$), using observed data. There are standard methodologies for estimation (say, the maximum likelihood estimator, MLE), or testing (such as the likelihood ratio test) that one may try to use. Of course, it still requires a great deal of effort to analytically compute these statistical indices and their (approximate) distributions on specific manifolds. A reasonably comprehensive account of these for the shape spaces of Kendall, or similar manifolds, may be found in Dryden and Mardia (1998).

The focus of the present article is a model independent, or nonparametric, methodology for inference on general manifolds. As a motivation consider the problem of discriminating between two distributions on an Euclidean space based on independent samples from them. In parametric inference one would use a density (with respect to a sigma-finite measure) which is specified except for a finite dimensional parameter as described above. One may use one of a number of standard asymptotically efficient procedures to test if the two distributions have different parameter values (See, e.g., Hotelling (1931), Goodall (1991)). If the statistician is not confident about this parametric model, or any other, one popular method is to test for the differences between the means of the two distributions by using the two sample means. When the sample sizes are reasonably large then the difference between the sample means is asymptotically normal with mean given by the difference between the population means. If the observations are from a normal distribution with the mean as the unknown parameter then this test is optimal in an appropriate sense (Bhattacharya et al. (2016), pp 296-300, Lehmann (1959), pp. 93,94). But used in other parametric model the test is not, in general, optimal and may even be inconsistent; that is, there may be many pairs of distributions $Q_1 \neq Q_2$ whose means are the same. However, when the components or coordinates of the distributions are such that the differences between $Q_1$ and $Q_2$ are reasonably expected to manifest in shifts of the mean vector, this widely used nonparametric test is quite effective, especially since with large sample sizes the asymptotic distribution is normal. Turning now to distributions $Q$ on non-Euclidean metric spaces $S$, one has an analogue of the mean given by the minimizer, if unique, of the average (with respect to $Q$) of the squared distance from a point. This is the so called \textit{Fréchet mean} introduced by Fréchet (1948), although physicists probably had used the notion earlier in specific physical contexts for the distribution $Q$ of the mass of a body, calling it the center of mass. Of course it is in general a non-trivial matter to find out broad conditions for the \textit{uniqueness of the Fréchet minimizer} and, in the case of uniqueness, to derive the (asymptotic) \textit{distribution of the sample Fréchet mean}. These allow one to obtain proper confidence regions for the Fréchet mean of $Q$ and critical regions for tests for detecting differences in means of distributions on $M$ (Bhattacharya and Patrangenaru, 2002, 2003, 2005). The theory of Fréchet means is presented in Section 2 (uniqueness and consistency), and in Section 4 (asymptotic distributions). The main results in Sections 2 and 4 are presented with complete proofs. Section 4 plays a central role for inference in the present context, and it contains some improvements of earlier results.
It has been shown in data examples that the nonparametric procedures based on Fréchet means often greatly outperform their parametric counterparts (See Bhattacharya and Bhattacharya (2012)). Misspecification of the model is a serious issue with parametric inference, especially for distributions on rather complex non-Euclidean spaces.

In this article two types of images and their analysis are distinguished. The greater emphasis is on landmarks based shapes introduced by Kendall (1984) and Bookstein (1991). This looks at a $k$-ad or a set of $k$ properly chosen points, not all the same, on an $m$-dimensional image (usually $m = 2$ or $3$), $k > m$, such as an MRI scan of a section of the brain for purposes of diagnosing a disease, or a scan of some organ of a species for purposes of morphometrics. In order to properly compare images taken from different distances and angles using perhaps different machines, the shape of a $k$-ad is defined modulo translation, scaling and rotation. The resulting shapes comprise Kendall’s shape spaces. In addition, one may consider affine shapes which are invariant under all affine transformations appropriate in scene recognition; similarly, projective shapes invariant under projective transformations are often used for robotic vision. The precise mathematical (geometric) descriptions of these kind of images are presented in Section 3. Sections 5 and 6 provide the asymptotic theory of tests and confidence regions on manifolds, based on the asymptotic distribution theory developed in Section 4.

Section 8 considers briefly the second type of images, namely, the actual geometric shape of a compact two-dimensional surface or a three dimensional body. Here the shape space is infinite dimensional and may be viewed as a Hilbert manifold (Ellingson et al., 2013). For purposes of diagnostics such as described above, this is probably not to be preferred in comparison with the finite dimensional landmarks based shapes considered by Kendall, because of the curse of dimensionality. The Hilbert manifolds then are better suited for purposes of machine vision. However, for that task a more effective methodology seems to be one which builds on the exciting inquiry of Kac (1966): Can one hear the shape of a drum? It turns out that for two-dimensional compact Riemannian manifolds such as compact surfaces, the spectrum of the Laplace Beltrami operator identifies the manifold in most cases, although there are exceptions. In three and higher dimensions, on the other hand, iso-spectral manifolds are not so rare (Milnor, 1964; Gordon et al., 1992; Zelditch, 2000). Still, computer scientists and other researchers in machine vision have successfully implemented algorithms to identify two and three-dimensional images by the spectrum of their Laplaceans, sometimes augmented by their eigen-functions (Demmel et al., 1999; Gotsman et al., 2003; Jain and Zhang, 2007; Shamir, 2006; Reuter et al., 2009). A mathematical breakthrough was achieved by Jones et al. (2008), who proved that indeed compact manifolds are determined by this augmentation.

Section 7 is devoted to another very important statistical problem: nonparametric classification via density estimation, and nonparametric regression on manifolds. In particular, we emphasize Ferguson’s nonparametric Bayes theory of using Dirichlet process priors for this endeavor (Ferguson, 1973, 1974).

Sections 9 provides a number of applications of the theory of Fréchet means, including Fisher’s example mentioned above, but focusing on two-sample problems on landmarks based shape spaces such as those introduced by Kendall (Kendall, 1984, 1989).

The appendix, Section 10, provides a ready access to some notions in Riemannian geometry used in the text.
2. Existence of the Fréchet Mean on Non-Euclidean Spaces.

Let $(S, \rho)$ be a metric space and $Q$ a probability measure on it. The Fréchet function of $Q$ is defined as

\[
F(p) = \int \rho^2(p, q)Q(dq), p \in S.
\]

If $F$ is finite at some $p$ then it is finite on $S$. The set $C(Q)$ of minimizers of $F$ is called the Fréchet mean set. If the minimizer is unique, i.e., $C(Q)$ is a singleton, then it is called the Fréchet mean of $Q$, and one says that the Fréchet mean of $Q$ exists. We will often use the topological condition

\[
\text{(2)} \quad \text{All closed bounded subsets of } S \text{ are compact.}
\]

When $S$ is a Riemannian manifold and $\rho = \rho_g$ is the geodesic distance on it, then (2) is equivalent to the completeness of $S$, by the Hopf-Rinow theorem (Do Carmo (1992), pp. 146-149).

Let $X_1, \ldots, X_n$ be a random sample from $Q$, i.e., $X_j$ are i.i.d. with common distribution $Q$, defined on a probability space $(\Omega, F, P)$. Denote by $F_n$ the Fréchet function of the empirical $Q_n = (1/n)\sum_{i=1}^n \delta_{X_i}$, where $\delta_x$ is the point mass at $x$. Also let $B^\epsilon = \{p \in S : \rho(p, B) < \epsilon\}$ for $B \subset S$.

**Theorem 2.1** (Bhattacharya and Patrangenaru, 2003). Assume (2) and that the Fréchet function $F$ of $Q$ is finite. Then (a) $C(Q)$ is nonempty and compact, and (b) for each $\epsilon > 0$, there exists a random positive integer $N = N(\omega; \epsilon)$ and a $P$-null set $\Gamma$ such that for all $n \geq N(\omega; \epsilon)$,

\[
C(Q_n) \subset (C(Q))^\epsilon \quad \text{for every } \omega \notin \Gamma.
\]

(c) In particular, if the Fréchet mean of $Q$, say $\mu$, exists, then every measurable selection $\mu_n$ from $C(Q_n)$, converges almost surely to $\mu$. In this case $\mu_n$ is called the sample Fréchet mean.

**Proof.** First assume $S$ is compact. Then (a) is obvious. To prove (b), it is enough to show that $\delta_n = \max\{|F_n(p) - F(p) : p \in S\} \to 0$ almost surely as $n \to \infty$. To see this let $\lambda = \min\{F(p) : p \in S\} = F(q) \forall q \in C(Q)$. If $(C(Q))^\epsilon = S$, then (3) holds with $N = 1$ (for every $\omega$). Assume $(C(Q))^\epsilon$ is not $S$, and write $M_1 = S \setminus (C(Q))^\epsilon$. There exists $\theta(\epsilon) > 0$, such that $\min\{F(p) : p \in M_1\} = \lambda + \theta(\epsilon)$. Also, there exists $\epsilon_1 > 0$, $\epsilon_1 \leq \epsilon$, such that $F(p) \leq \lambda + \theta(\epsilon)/4 \forall p \in (C(Q))^\epsilon_1$. Since $\delta_n \to 0$ a.s., there exists $N = N(\omega)$ such that such that $\forall n \geq N$, $F_n(p) < \lambda + \theta(\epsilon)/3 \forall p \in (C(Q))^\epsilon_1$ and $F_n(p) > \lambda + \theta(\epsilon)/2 \forall p \in M_1$, so that $C(Q_n) \subset (C(Q))^\epsilon$, proving (3). In order to show that $\delta_n \to 0$ a.s. first note that, irrespective of $Q$, $|F(p) - F(p')| \leq c\rho(p, p')$ where $c = 2\max\{\rho(q, q') : q, q' \in S\}$. Given any $\delta > 0$, $|F(p) - F(p')| < \delta/4$ if $\rho(p, p') < \eta = \delta/4c$. Let $q_1, \ldots, q_k$ be such that the balls $B(q_i : \eta)$ with radius $\eta$ and center $q_i$ cover $S$. Then $|F(p) - F(q_i)| < \delta/4 \forall p \in B(q_i : \eta)(i = 1, \ldots, k)$. The same is true with $Q$ replaced by $Q_n$. By the strong law of large numbers (SLLN), there exists $N_1 = N_1(\omega; \delta)$ such that $|F_n(q_i) - F(q_i)| < \delta/2 \forall n \geq N_1 (i = 1, \ldots, k)$, outside a $P$-null set. It follows that, outside a $P$-null set, $|F_n(p) - F(p)| < |F_n(p) - F_n(q_i)| + |F_n(q_i) - F(q_i)| + |F(q_i) - F(p)| < \delta \forall p \in B(q_i : \eta)(i = 1, \ldots, k)$, provided $n \geq N_1$.

Consider now the non-compact case, but assuming (2). Let $\lambda = \inf\{F(p) : p \in S\}$. This infimum is attained in $S$. To see this, let $p_k (k = 1, 2, \ldots)$ be such that $F(p_k) \to \lambda$ as
$k \to \infty$. Since $\rho(p, q) \leq \rho(p, x) + \rho(q, x) \forall p, q, x$, one has

\begin{equation}
\rho(p, q) \leq \int \rho(p, x)Q(dx) + \int \rho(q, x)Q(dx) \leq F^{1/2}(p) + F^{1/2}(q), \forall p, q \in S.
\end{equation}

Letting $p = p_k$ and $q = p_k$, one obtains $\lim \sup \rho(p_k, p_1) < \infty$. Hence the sequence $\{p_k\}$ is bounded, and its closure is compact. Therefore, there exists $p^*$ such that $F(p^*) = \lambda$. Thus $C(Q)$ is nonempty and closed. If $q$ is any point in $C(Q)$ then taking $p = p^*$ and $q \in C(Q)$ in (4), one has $\rho(p^*, q) \leq 2\lambda^{1/2}$. That is $C(Q) \subset B(p^*, \lambda^{1/2})$. Thus part (a) is proved. To prove part (b), one has, using $Q_n$ for $Q$ and a fixed point $p^*$ for $q$ in $C(Q)$ in (4), the inequality $F_n^{1/2}(p) \geq \rho(p, p^*) - F_n^{1/2}(p^*), \forall p$. Fix a $\delta > 0$. Consider the compact set $M_1 = \{q : \rho(q, p^*) \leq 2(\lambda + \delta)^{1/2} + \lambda^{1/2}\}$. Then for $p \in S \setminus M_1$, one has $F_n(p) \geq [2(\lambda + \delta)^{1/2} + \lambda^{1/2} - F_n^{1/2}(p^*)]^2 > \lambda + \delta$. Hence $F_n(p^*) < \lambda + \delta$ for all sufficiently large $n \geq N_1 = N_1(\omega)$ except for $\omega$ lying in a $P$-null set, in view of the SLLN. Hence $C(Q_n) \subset M_1$ for $n \geq N_1$. Applying the result in the compact case (with $S = M_1$), one arrives at (b). Part (c) is an immediate consequence of part (b).

For compact metric spaces $S$, part (c) of Theorem 2.1 follows from Ziezold (1977).

**Remark 2.2.** Theorem 2.1 extends to more general Fréchet functions, including $F(p) = \int \rho^\alpha(p, q)Q(dq), \alpha \geq 1$.

**Remark 2.3.** Relation (3) does not imply that the sets $C(Q)$ and $C(Q_n)$ are asymptotically close in the Hausdorff distance. Indeed, in many examples $C(Q_n)$ may be a singleton, while $C(Q)$ is not. See, e.g., Bhattacharya and Patrangenaru (2003), Remark 2.6, where it is shown that whatever be the absolutely continuous distribution $Q$ on $S^1$, $C(Q_n)$ is almost surely a singleton; in particular, this is the case when $Q$ is the uniform distribution for which $C(Q) = S^1$. In view of this, and for asymptotic distribution theory considered later, it is important to find broad conditions on $Q$ for the existence of the Fréchet mean (as the unique minimizer of the Fréchet function).

Let $S = M$ be a differentiable manifold of dimension $d$—a topological space which is metrizable as a separable metric space such that (i) every $p \in M$ has an open neighborhood up with a homeomorphism $\psi_p : U_p \to B_p$, where $B_p$ is an open subset of $\mathbb{R}^d$, and (ii) (compatibility condition) if $U_p \cap U_q$ is nonempty, then the map $\psi_p \circ \psi_q^{-1} : \psi_q(U_p \cap U_q) \to \psi_p(U_p \cap U_q)$ is a $C^\infty$-a common example is the sphere $S^d = \{x \in \mathbb{R}^{d+1} : |x| = 1\}$; one may take $p$ as the north pole $(0,0,\ldots,0,1)$ and $q$ as the south pole $(0,0,\ldots,0,-1)$, $U_p = S^d \setminus \{q\}, U_q = S^d \setminus \{p\}$, and $\psi_p$ and $\psi_q$ are the stereographic projections on $S^d \setminus \{q\}$ and $S^d \setminus \{p\}$, respectively, onto $\mathbb{R}^d$. Or, one may take $2d$ open hemispheres $U_p$ of $S^d$ with poles whose coordinates are all zeros, except for $+1$ or $-1$ at the $i$-th coordinate ($i = 1, \ldots, d$), each mapped diffeomorphically onto the open unit disc in $\mathbb{R}^d$. There are infinitely many distances which metrize the topology of $M$. The two most common are (1) the Euclidean distance under an embedding, and (2) the geodesic distance when $M$ is endowed with a metric tensor. For the first, recall that a smooth $(C^\infty)$ map $J : M \to E^N$ is an embedding into an Euclidean space $E^N$, if (a) $J$ is one-to-one and $M \to J(M)$ is a homeomorphism with $J(M)$ given the relative topology of $E^N$, and (b) the differential $d_pJ$ on the tangent space $T_p(M)$ into the tangent space of $E^N$ at $J(p)$ is one-to-one. The Euclidean distance on $J(M)$ (transferred to $M$ via $J^{-1}$) is called the extrinsic distance $\rho_J$ on $M$. The embedding is said to be closed if $J(M)$ is closed. For $S^d$ one may, for example, take $J$ to be the inclusion map of $S^d$ into $\mathbb{R}^{d+1}$, and the extrinsic distance is the chord distance.
Theorem 2.4 ((Bhattacharya and Patrangenaru, 2005) (Extrinsic Fréchet Mean on a Manifold)). Let $M$ be a differentiable manifold and $Q$ a probability measure on it. If $J$ is a closed embedding of $M$ into an Euclidean space $E^N$, and the Fréchet function of $Q$ is finite with respect to the induced Euclidean distance on $J(M)$, then the (extrinsic) Fréchet mean exists as the unique minimizer of the Fréchet function if and only if there is a unique point $\bar{\mu}_{J,E}$ in $J(M)$ closest to the Euclidean mean $m$ of the (push forward) distribution $Q_J = Q \circ J^{-1}$ on $E^N$, and then the extrinsic mean is $J^{-1}\bar{\mu}_{J,E}$.

Proof. For a point $c \in J(M)$, writing $|y|^2 = \sum_{i=1}^N (y^{(i)})^2$ for the usual squared Euclidean norm on $E^N$,

$$
\int_{J(M)} |c - y|^2 Q_J(dy) = \int_{E^N} |c - y|^2 Q_J(dy) = \int_{E^N} |m - y|^2 Q_J(dy) + |c - m|^2.
$$

This is minimized with respect to $c$, by setting $c$ to be the point in $J(M)$ closest to $m$, if there is only one such point, and the minimizer is not unique otherwise. 

Example 2.5 (Extrinsic Mean on the Sphere $S^d$). Let the inclusion map on $S^d$ into $\mathbb{R}^{d+1}$ be the embedding $J$. Then the mean $m$ of $Q_J$ on $\mathbb{R}^{d+1}$ lies inside the unit ball $B(0:1)$ in $\mathbb{R}^{d+1}$ unless $Q$ is degenerate at a point $m \in S^d$. If $Q$ is nondegenerate, the closest point to $m$ in $S^d$ is $m/|m|$ unless $m = 0$ (i.e., $m$ lies at the center of the unit ball). Thus (the image of) the extrinsic mean is $\mu_{J,E} = m/|m|$. If $m = 0$, then $C(Q) = S^d$. If $Q$ is degenerate at $m$, then $m$ is the extrinsic mean. Taking $Q$ to be the empirical $Q_n$, the sample Fréchet mean is $\bar{X}/|\bar{X}|$, if $\bar{X}$ is not the origin in $\mathbb{R}^{d+1}$. If $\bar{X} = 0$, then $C(Q_n) = S^d$.

Theorem 2.4 allows one in many important cases of interest in image analysis to find analytic characterizations for the existence of the extrinsic mean (i.e., as the unique minimizer of the Fréchet function) and computable formulas for its computation. This will be discussed in Section 3.

Unfortunately, on a Riemannian manifolds $M$ with metric tensor $g$ there is no good analog of Theorem 2.4 for the intrinsic mean of $Q$, – the minimizer of the Fréchet function under the geodesic distance $\rho_g$. The pioneering work by Karcher (1977) followed by generalizations and strengthening, most notably, by Kendall (1990), Le (2001) and Afsari (2011) hold under support restrictions on $Q$, which are untenable for general statistical inference. The recent results of Afsari (2011) are the sharpest among these, which we state below (for the Fréchet function (1)) without proof. For the terminology used in the statement we refer to the Appendix on Riemannian geometry. Recall that the support of a probability measure $Q$ on a metric space is the smallest closed set $D$ such that $Q(D) = 1$.

Theorem 2.6 (Afsari, 2011) (Intrinsic Mean on a Riemannian Manifold)). On a complete Riemannian manifold $(M, g)$, there exists an intrinsic Fréchet mean of $Q$, as the unique minimizer of the Frechet function (1) with the geodesic distance $\rho = \rho_g$, if the support of $Q$ is contained in a geodesic ball of radius less than $r^* = (1/2) \min\{\text{inj}(M), \pi/\sqrt{C}\}$. Here $\text{inj}(M)$ is the injectivity radius of $M$; and $C$ is the supremum of sectional curvatures of $M$, if positive, or zero otherwise.

Remark 2.7. If the Riemannian manifold $M$ is complete, simply connected and has nonpositive curvature and the Fréchet function of $Q$ is finite, then the intrinsic mean of $Q$ exists (as the unique minimizer of $F$). An important generalization of this is to the so called metric spaces of non-positive curvature, or the NPC spaces, which include many interesting metric spaces which are not manifolds. Such spaces were introduced by Alexandrov (1957)
and further developed by Reshetnyak (1968) and Gromov (1981). See Sturm (2003) for a fine exposition.

**Example 2.8.** Let $M = S^2$. Then it has constant sectional curvature 1, and its injectivity radius is $\pi$. Thus if $Q$ has support contained in an open hemisphere, then the Fréchet mean of $Q$ under the geodesic distance exists. To see that one cannot relax this support condition in general, consider the uniform distribution on the equator. Then the minimum expected squared distance is attained at both the North and South poles (say, $(0,0,1)$, and $(0,0, -1)$), so that $C(Q)$ has two points.

**Remark 2.9.** For purposes of statistical inference the support condition in Theorem 2.6 is restrictive, but as Example 2.8 shows one cannot dispense with the support condition without some further conditions on the nature of $Q$. In statistical practice a reasonable assumption is that the distribution is absolutely continuous. In $S^3$ under the assumption that $Q$ has a continuous density (with respect to the arc length measure on intervals, i.e., the Lebesgue measure on $[0, 2\pi]$) a necessary and sufficient condition, which applies broadly, was obtained in Bhattacharya (2007) and may be found in Bhattacharya and Bhattacharya (2012), pp. 31-33, 73-75.

3. Geometry of Kendall’s Shape Spaces.

3.1. Kendall’s Similarity Shape Space $\Sigma^k_m$. The similarity shape of a $k$-ad $x = (x_1, \ldots, x_k)$ in $\mathbb{R}^m$, not all points the same, is its orbit under the group generated by translations, scaling and rotations. Writing $\bar{x} = (x_1 + \cdots + x_k)/k, \langle \bar{x}, \cdots, \bar{x} \rangle$, the effect of translation is removed by looking at $(x_1 - \bar{x}, \ldots, x_k - \bar{x}) = x - \langle \bar{x} \rangle$, which lies in the $mk - m$ dimensional hyperplane $L$ of $\mathbb{R}^{mk}$ made up of $m \times k$ matrices with the $m$ row sums all equal to zero. To get rid of scale, one looks at $u = (x - \langle \bar{x} \rangle)/\|x - \langle \bar{x} \rangle\|$, where $\|\cdot\|$ is the usual norm in $\mathbb{R}^{mk}$. This translated and scaled $k$-ad is called the preshape of the $k$-ad. It lies on the unit sphere in $L$, and is isomorphic to $S^{m(k-1)-1}$. An alternative representation of the preshape, which we use, is obtained as $p = xH/\|xH\|$, where $H$ is the $k \times (k - 1)$ Helmert matrix comprising $k - 1$ column vectors forming an orthonormal basis of $1^k$, namely, the subspace of $\mathbb{R}^k$ orthogonal to $(1, \ldots, 1)'$. A standard $H$ has the $j$-th column given by $(a(j), \ldots, a(j), -ja(j), 0, \ldots, 0)'$, where the first $j$ elements equal $a(j) = \lfloor j(j+1)/2 \rfloor$ $(j = 1, \ldots, k - 1)$. Then $p$ is an $m \times (k - 1)$ matrix of norm one. The shape $\sigma(x) = \sigma(p)$ of $x$ is then identified with the orbit of $p$ under all rotations:

$$\sigma(x) = \sigma(p) = \{Ap : A \in SO(m)\},$$

where $SO(m) = \{A : AA' = I_m, det(A) = 1\}$ is called the special orthogonal group acting on $\mathbb{R}^m$. The set of all shapes $\sigma(p)$ is Kendall’s similarity shape space $\Sigma^k_m$.

If $m = 2, k > 2$, the action of $SO(2)$ on the preshape sphere $S^{2k-3}$ is free, i.e., no $A \in SO(2)$ other than the identity has a fixed point and each orbit of a point in $S^{2k-3}$ has an orbit of dimension one, namely the dimension of $SO(2)$. Since each $A \in SO(2)$ is an isometry of $S^{2k-3}$ endowed with the geodesic distance, it follows that $\Sigma^k_2 = S^{2k-3}/SO(2)$ is a Riemannian manifold. For $m > 2, k > m$, however, the action of $SO(m)$ on $S^{m(k-1)-1}$ is not free. For example, for $m = 3$, each collinear $k$-ad in $SO(3(k-1)-1)$ is invariant under all rotations in $\mathbb{R}^3$ around the line of the $k$-ad. $\Sigma^k_3$ is then a disjoint union of two Riemannian manifolds, not complete, one comprising of the orbits of collinear $k$-ads under rotation by elements of $SO(2)$ other than those that keep it fixed (except for the identity). The other comprises of orbits under $SO(3)$ of all non-collinear $k$-ads in $S^{3(k-1)-1}$. $\Sigma^k_3$ is then a stratified space with two strata. More generally, $\Sigma^k_m, m > 2 (k > m)$, is a stratified space.
with $m - 1$ strata. See Kendall et al. (1999), Chapter 6, for a complete description of the intrinsic geometry of $\Sigma^k_m$. Also see Huckemann et al. (2010) for intrinsic analysis of more general stratified spaces of the form $M = N/G$, where $N$ is a Riemannian manifold and $G$ is a Lie group of isometries acting on $N$.

3.1(a) **Intrinsic geometry of $\Sigma^k_2$.** For the case $m = 2$, it is convenient to regard a $k$-ad $x = ((x_1, y_1), \cdots, (x_k, y_k))$ as a $k$-tuple $z = (z_1, \cdots, z_k)$ of numbers $z_1 = x_1 + i y_1, \cdots, z_k = x_k + i y_k$ in the complex plane $\mathbb{C}$, and let $p = (z - <z>) / |z - <z>|$. Then the shape of $p$, or $z$, is identified with the orbit

$$\sigma(z) = \sigma(p) = \{e^{i\theta}p : \theta \in (-\pi, \pi]\}. \tag{7}$$

One may equivalently, consider the shape as the orbit $\{\lambda(z - <z>) : \lambda \in \mathbb{C}\}$. That is, after Hérmertization, the shape of $x$, or $z$, is identified with a complex line passing through the origin in $\mathbb{C}^{k-1}$. The shape space is then identified with the complex projective space $\mathbb{C}P^{k-2}$, of (real) dimension $2k - 4$. We will, however, use the representation $\Sigma^k_2 = \mathbb{C}S^{k-1}/G$, where $G = \{e^{i\theta} : \theta \in (-\pi, \pi]\}$ is a 1-dimensional compact group ($G \simeq S^1$) of isometries of the preshape sphere $\mathbb{C}S^{k-1} = \{p = (p_1, \cdots, p_{k-1}) \in \mathbb{C}^{k-1} : |p| = 1\}$, which is isomorphic to $S^{2k-3}$. Recall that the metric tensor on $S^{2k-3} \simeq \mathbb{C}S^{k-1}$ is that inherited from the inclusion map into $\mathbb{R}^{2(k-1)} = \{(x_1, y_1, x_2, y_2, \cdots, x_{k-1}, y_{k-1}) : (x_j, y_j) \in \mathbb{R}^2 \forall j\} \simeq \mathbb{C}^{k-1} = \{(z_1, z_2, \cdots, z_{k-1}) : z_j = x_j + i y_j \in \mathbb{C} \forall j\}$. That is, the inner product at the tangent space $T_p\mathbb{C}S^k$ is $\langle \tilde{v}, \tilde{w} \rangle = \text{Re}(\tilde{v}^\ast \tilde{w}^\ast)$, when $\tilde{v}, \tilde{w}$ are expressed as complex $1 \times (k-1)$ matrices (row vectors) in $\mathbb{C}S^{k-1}$, satisfying $\text{Re}(\tilde{v}^\ast \tilde{w}^\ast) = 0 = \text{Re}(\tilde{w}^\ast \tilde{v}^\ast)$. The projection map is then $\pi : p \rightarrow \sigma(p)$. The **vertical subspace** $V_p$ is obtained by differentiating the curve $\theta \rightarrow e^{i\theta}p$, say at $\theta = 0$, yielding $i p$. That is, $V_p = \{cip : c \in \mathbb{R}\}$. Thus the **horizontal subspace** is $H_p = \{\tilde{v} : \text{Re}(\tilde{v}^\ast \tilde{w}^\ast) = 0, \text{Re}(i\tilde{v}^\ast \tilde{w}^\ast) = 0\} = \{\tilde{v} : \tilde{v}^\ast = 0\}$. The geodesics $\gamma(t; \sigma(p), v)$ for $v = (d_p, \pi)\tilde{v}$ (for $\tilde{v} \in H_p$), and the exponential map $\text{Exp}_{\sigma(p)}(p, \Sigma^k_2)$ are specified by this isometry between $T_{\sigma(p)}(\Sigma^k_2)$ and $H_p$ for all shapes $\sigma(p)$ (See the Appendix, Section A). Thus, identifying vectors $v$ in $H_p$ with vectors $v$ in $T_{\sigma(p)}(\Sigma^k_2)$, one obtains

$$T_{\sigma(p)}(\Sigma^k_2) = \{v = (d_p, \pi)\tilde{v} : \forall v \text{ such that } p\tilde{v}^\ast = 0\} \tag{8}$$

$$\text{Exp}_{\sigma(p)}(0, \sigma(p), v) = \sigma(p) \cdot \text{Exp}_{\sigma(p)}(0, \sigma(p), v) = \sigma(\cos(\langle \hat{v} | p \rangle \cdot \sin(\langle \hat{v} | \hat{v} \rangle / |\hat{v}|) \cdot \langle v \neq 0, p\tilde{v}^\ast = 0\rangle; \gamma(t; \sigma(p), v) = \sigma((\cos t|\hat{v}|p + \sin(t|\hat{v}|\hat{v}) / |\hat{v}|), (t \in \mathbb{R}, p\tilde{v}^\ast = 0), v \neq 0).$$

Denoting by $\rho_{gs}$ and $\rho_g$ the geodesic distances on $\mathbb{C}S^{k-1}$ and $\Sigma^k_2$, respectively, and recalling that (See Example 10.1 and Kendall et al. (1999), p.114) $\rho_{gs}(p, q) = \text{arccos}(\text{Re}pq^\ast)$, one has

$$\rho_g(\sigma(p), \sigma(q)) = \inf\{\rho_{gs}(p, q) : p \in O_w, q \in O_w\} \tag{9}$$

$$= \inf\{\text{arccos}(\text{Re}(pq^\ast e^{i\theta})^\ast) : \theta \in [0, 2\pi]\}$$

$$= \text{arccos}|(pq^\ast e^{i\theta})| (1 - |\text{Re}(pq^\ast e^{i\theta})|)^{2}^{1/2} (qe^{i\theta} - (pq^\ast e^{i\theta})p),$$

It follows that the geodesics are periodic with period $\pi$, and the cut locus of $\sigma(p)$ is $\{\sigma(q) : all q such that \text{arccos}(|pq^\ast|) = \pi/2\}$, and that the injectivity radius of $\Sigma^k_2$ is $\pi/2$. The inverse exponential map is given by $\text{Exp}_{\sigma(p)}^{-1}(\sigma(q)) = v$, where $v = (d_p, \pi)\tilde{v}$ ($\tilde{v} \in H_p$), and $\tilde{v}$ satisfies (Use (A.3) with the representation of $S^{2k-3}$ as $\mathbb{C}S^{k-1}$)

$$\tilde{v} = \text{Exp}_{\sigma(p)}^{-1}(qe^{i\theta}) \tag{10}$$

$$= \text{arccos}(\text{Re}(pq^\ast e^{i\theta})) (1 - |\text{Re}(pq^\ast e^{i\theta})|^2)^{1/2} (qe^{i\theta} - (pq^\ast e^{i\theta})p),$$
where \( \theta \) is so chosen as to minimize \( \rho_{gs}(p, q e^{i\theta}) = \arccos(Re(pq^* e^{-i\theta})) \). That is, \( (pq^* e^{-i\theta}) = |pq^*| \), or \( e^{i\theta} = pq^*/|pq^*| \) (for \( pq^* \neq 0 \), i.e., for \( \sigma(q) \) not in \( C(\sigma(p)) \). Hence, writing \( \rho = \rho_g(\sigma(p), \sigma(q)) \), \( \rho \neq 0 \), one has

\[
\tilde{v} = \sqrt{1 - |pq^*|^2} \left( |pq^*|^2 (pq^*/|pq^*|)q - pq^* |p\right) \] (11)

\[
= \left[ p/\sin \rho \right] (q e^{i\theta} - \cos \rho p) \]

This horizontal vector \( \tilde{v} (\in H_p) \) represents \( \text{Exp}^{-1}(\sigma(q)) = v \).

The sectional curvature of \( \Sigma_{A}^{k} \) at a section generated by two orthonormal vector fields \( \tilde{W}_1 \) and \( \tilde{W}_2 \) is \( 1 + 3 \cos^2 \phi \) where \( \cos \phi = \langle U_1, iU_2 \rangle \), \( U_1 \) and \( U_2 \) being the horizontal lifts of \( W_1 \) and \( W_2 \) (See Do Carmo (1992)).

3.1(b). **Extrinsic geometry of \( \Sigma_{A}^{k} \) induced by an equivariant embedding.** As mentioned in Section 2, no broad sufficient condition is known for the existence of the intrinsic mean (i.e., of the uniqueness of the minimize of the corresponding Fréchet function). The extrinsic mean, on the other hand, is unique for most \( Q \), and is generally computable analytically. However, for an extrinsic analysis to be very effective one should choose a good embedding which retains as many geometrical features of the shape manifold as possible. Let \( \Gamma \) be a Lie group acting on a differentiable manifold \( M \), and denote by \( GL(N, \mathbb{R}) \) the general linear group of nonsingular transformations on a Euclidean space \( \mathbb{E}^N \) of dimension \( N \) onto itself. An embedding \( J \) on \( M \) into \( \mathbb{E}^N \) is said to be \( \Gamma \)-equivariant if there exists a group homomorphism \( \Phi : \gamma \rightarrow \phi_\gamma \) of \( \Gamma \) into \( GL(N, \mathbb{R}) \) such that \( J(\gamma p) = \phi_\gamma(Jp) \) \( \forall p \in M, \gamma \in \Gamma \). Often, when there is a natural Riemannian structure on \( M, \Gamma \) is a group of isometries of \( M \). Consider the so-called Veronese-Whitney embedding \( J \) of \( \Sigma_{A}^{k} \) into the (real) vector space \( S(k-1, \mathbb{C}) \) of all \( (k-1) \times (k-1) \) Hermitian matrices \( B = B^* \), defined by

\[
J\sigma(p) = p^*p \quad [\sigma(p) = \{ e^{i\theta} p, \theta \in [0, 2\pi), p \in \mathbb{C}S^{k-1} \}] \]

The Euclidean inner product on \( S(k-1, \mathbb{C}) \), considered as a real vector space, is given by \( \langle B, C \rangle = \text{Re} (\text{Trace}(BC^*)) \). Let \( SU(k-1) \) denote the special unitary group of all \( (k-1) \times (k-1) \) unitary matrices \( A \) (i.e., \( A^*A = I, \det(A) = 1 \)) acting on \( S(k-1, \mathbb{C}) \) by \( B \rightarrow A^*BA \). Then the embedding (12) is \( \Gamma \)-equivariant, with \( \Gamma = \{ \gamma_A: A \in SU(k-1) \} \) and the group action on \( \Sigma_{A}^{k} \) given by: \( \gamma_A \sigma(p) = \sigma(pA) \). For \( J\sigma(pA) = A^*p^*Ap = \phi(\gamma_A)(J\sigma(p)) \), say, where the group homomorphism on \( \Gamma \) onto \( SU(k-1) \) is given by \( \gamma_A \rightarrow \phi(\gamma_A) : \phi(\gamma_A)B = A^*BA \). Note that \( SU(k-1) \) is a group of isometries of \( S(k-1, \mathbb{C}) \). In the notation defining equivariance, one lets \( S(k-1, \mathbb{C}) = \mathbb{E}^N (N = (k-1)^2), SU(k-1) \) is a subgroup of \( GL(N, \mathbb{R}) \).

To compute the extrinsic mean of \( Q \) on \( \Sigma_{A}^{k} \), let \( QJ = Q \circ J^{-1} \) be the probability induced on \( S(k-1, \mathbb{C}) \) by the map \( J \) in (12), and let \( \mu_J \) denote its Euclidean mean. By Theorem 2.1, the (image of the) extrinsic mean of \( Q \) is given by the orthogonal projection \( P \) on \( J(\Sigma_{A}^{k}) \).

**Proposition 3.1.** (Bhattacharya and Patrangenaru, 2003) The image under \( J \) of the extrinsic mean of \( Q \) comprises all elements of the form \( w^*w \) where \( w^* \) is a normalized (column) eigenvector with the largest eigenvalues of \( \mu_J \). In particular, the extrinsic mean of \( Q \) exists if and only if the largest eigenvalue of \( \mu_J \) is simple.

**Proof.** Let \( T \) be a \( (k-1) \times (k-1) \) unitary matrix such that \( T\mu_J T^* = D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{k-1}) \) where \( \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{k-1} \) are the ordered eigenvalues of \( \mu_J \). Then the columns of \( T^* \) form a complete orthonormal set of eigenvectors of \( \mu_J \). By relabelling the landmarks, if necessary, we may assume that the \( i \)th column of \( T^* \) is an eigenvector with eigenvalue \( \lambda_i \).
Write \( \|A\|^2 = \text{Trace}(AA^*) \) as the square of the Euclidean norm of \( A \). Then for elements \( w^*w \) of \( J(\Sigma_k) = \{ w^*w, w \in \mathbb{C}S^{k-1} \} \), denoting \( v = wT^* \), one has
\[
\|w^*w - \mu_j\|^2 = \|T w^*wT^* - T \mu_j T^*\| = \|v^*v - D\|^2 = \sum_{i,j} |\bar{v}_i v_j - \lambda_i \delta_{ij}|^2 = 1 + \sum_i \lambda_i^2 - 2 \sum_i \lambda_i |v_i|^2
\]
which is minimized over \( J(\Sigma_2) \) by taking \( v_{k-1} = 1 \) and \( v_i = 0 \forall i < k - 1 \), i.e., by taking \( w^* = T*v^* \) be any normalized eigenvector of \( \mu_j \) with the largest eigenvalue. \( \square \)

A size-and-shape similarity shape \( s\sigma(z) \) is defined for Helmertized \( k \)-ads \( z = (z_1, \cdots, z_{k-1}) \) as its orbit under \( SO(m) \). An equivariant embedding for it is \( s\sigma(z) \rightarrow z^*z/|z| \), on the size-and-shape-similarity shape space \( S\Sigma_k^1 \) into \( S(k - 1, \mathbb{C}) \).

### 3.2. Reflection Similarity Shape Space \( \tilde{R}\Sigma_m^k \), \( m > 2, k > m \)
For \( m > 2 \), let \( \tilde{N}S^m(k-1)^{-1} \) be the subset of the centered preshape sphere \( S^m(k-1)^{-1} \) whose points \( p \) span \( \mathbb{R}^m \), i.e., which, as \( m \times k \) matrices, are of full rank. We define the reflection similarity shape of the \( k \)-ad as
\[
(13) \quad r\sigma(p) = \{ Ap : A \in O(m) \} \quad (p \in \tilde{N}S^m(k-1)^{-1}),
\]
where \( O(m) \) is the set of all \( m \times m \) orthogonal matrices \( A : AA^T = I_m, \det(A) = \pm 1 \). The set \( \{ r\sigma(p) : p \in \tilde{N}S^m(k-1)^{-1} \} \) is the reflection similarity shape space \( \tilde{R}\Sigma_m^k = \tilde{N}S^m(k-1)^{-1} / O(m) \). Since \( \tilde{N}S^m(k-1)^{-1} \) is an open subset of the sphere \( S^m(k-1)^{-1} \), it is a Riemannian manifold. Also \( O(m) \) is a compact Lie group of isometries acting on \( S^m(k-1)^{-1} \). Hence there is a unique Riemannian structure on \( \tilde{R}\Sigma_m^k \) such that the projection map \( p \rightarrow r\sigma(p) \) is a Riemannian submersion.

We next consider a useful embedding of \( \tilde{R}\Sigma_m^k \) into the vector space \( S(k, \mathbb{R}) \) of all \( k \times k \) real symmetric matrices (See Bandulasiri and Patrangenaru (2005), Bandulasiri et al. (2009), Dryden et al. (2008), and Bhattacharya (2008)). Define
\[
(14) \quad J(r\sigma(p)) = p'p \quad (p \in \tilde{N}S^m(k-1)^{-1}),
\]
with \( p \) an \( m \times (k - 1) \) matrix with norm one. Note that the right side is a function of \( r\sigma(p) \). Here the elements \( p \) of the preshape sphere are Helmertized. To see that this is an embedding, we first show that \( J \) is one-to-one on \( \tilde{R}\Sigma_m^k \) into \( S(k - 1, \mathbb{R}) \). For this note that if \( J(r\sigma(p)) \) and \( J(r\sigma(q)) \) are the same, then the Euclidean distance matrices \( \{(|p_i - q_j|)_{1 \leq i \leq k - 1} \} \) and \( \{(|q_i - q_j|)_{1 \leq i \leq k - 1} \} \) are equal. Since \( p \) and \( q \) are centered, by geometry this implies that \( q_i = Ap_i (i = 1, \cdots, k - 1) \) for some \( A \in O(m) \), i.e., \( r\sigma(p) = r\sigma(q) \). We omit the proof that the differential \( d_pJ \) is also one-to-one. It follows that the embedding is equivariant with respect to a group action isomorphic to \( O(k - 1) \).

**Proposition 3.2** (Bhattacharya, 2008). (a) The projection of \( \tilde{\mu} \) into \( J(\tilde{R}\Sigma_m^k) \) is given by
\[
(15) \quad P_{J(\tilde{R}\Sigma_m^k)}(\tilde{\mu}) = \{ A : A = \sum_{j=1}^m (\lambda_j - \bar{\lambda} + \frac{1}{m} U_j U_j^T) \}
\]
where \( \lambda_1 \geq \cdots \geq \lambda_k \) are the ordered eigenvalues of \( \tilde{\mu} \), \( U_1, \ldots, U_k \) are corresponding orthonormal (column) eigenvectors and \( \bar{\lambda} = \frac{\sum_{i=1}^m \lambda_i}{m} \). (b) The projection set is a singleton and \( Q \) has a unique extrinsic mean \( \mu_E \) iff \( \lambda_m > \lambda_{m+1} \). Then \( \mu_E = \sigma(F) \) where \( F = (F_1, \ldots, F_m)' \), \( F_j = \sqrt{\lambda_j - \bar{\lambda} + \frac{1}{m} U_j} \).
For a detailed proof see Bhattacharya (2008), or Bhattacharya and Bhattacharya (2012), pp. 114, 115.

For \( m > 2 \), a size-and-reflection shape \( sr\sigma(z) \) of a Helmertized \( k \)-ad \( z \) in \( \mathbb{R}^m \) of full rank \( m \) is given by its orbit under the group \( O(m) \). The space of all such shapes is the size-and-reflection shape space \( SR\Sigma^k_m \). An \( O(k-1) \)-equivariant embedding of \( SR\Sigma^k_m \) into \( S(k-1, \mathbb{R}) \) is: \( J(sr\sigma(z)) = z'|z| \).

### 3.3. Affine Shape Space \( A\Sigma^k_m \)

Let \( k > m + 1 \). Consider the set of all \( k \)-ads in \( \mathbb{R}^m \), with full rank \( m \) as \( m \times k \) matrices. The affine shape of a \( k \)-ad \( x \) may be identified with its orbit under all affine transformations:

\[
\sigma(x) = \{Ax + c : A \in GL(m, \mathbb{R}), c \text{ an } m \times k \text{ matrix}\}.
\]

If the \( k \)-ad is centered as \( u = x - \bar{x} \), then the affine shape of \( x \), or of \( u \), is given by

\[
\sigma(x) = \sigma(u) = \{Au : A \in GL(m, \mathbb{R})\}, \ (u \text{ centered } k \text{-ad of rank } m).
\]

The space of all such affine shapes is the affine shape space \( A\Sigma^k_m \). Note that two Helmertized \( k \)-ads \( u \) and \( v \) (as \( m \times (k-1) \) matrices of full rank) have the same shape if and only if the rows of \( u \) and \( v \) span the same \( m \)-dimensional subspace of \( \mathbb{R}^{k-1} \). Hence we can identify \( A\Sigma^k_m \) with the Grassmannian \( G_m(k-1) \), namely, the set of all \( m \)-dimensional subspaces of \( \mathbb{R}^{k-1} \) (Sparr, 1992). For the Grassmann manifold, refer to Boothby (1986), pp. 63, 168, 362, 363. For extrinsic analysis on \( A\Sigma^k_m \approx G_m(k-1) \), consider the embedding of \( A\Sigma^k_m \) into \( S(k-1, \mathbb{R}) \) given by

\[
J(\sigma(u)) = FF',
\]

where \( F = (f_1 \cdots f_m) \) is a \( (k-1) \times m \) matrix and \( \{f_1, \cdots, f_m\} \) is an orthonormal basis of the \( m \)-dimensional subspace \( L \), say, of \( \mathbb{R}^{k-1} \) spanned by the rows of \( u \). Note that the \( (k-1) \times (k-1) \) matrix \( FF' \) is idempotent and is the matrix of orthogonal projection of \( \mathbb{R}^{k-1} \) onto \( L \). It is independent of the orthonormal basis chosen. The embedding is \( O(k-1) \)-equivariant under the group action \( \sigma(u) \rightarrow \sigma(uO) \) \((O \in O(k-1))\) on \( A\Sigma^k_m \), with \( O(k-1) \) acting on \( S(k-1, \mathbb{R}) \) by \( A \rightarrow AO' \).

**Proposition 3.3.** (Sugathadasa, 2006) The projection of \( \bar{\mu} \) into \( J(A\Sigma^k_m) \) is given by

\[
P(\bar{\mu}) = \left\{ \sum_{j=1}^{m} U_j U'_j \right\}
\]

where \( U = (U_1, \ldots, U_k) \in SO(k) \) is such that \( \bar{\mu} = U\Lambda U' \), \( \Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_k) \), \( \lambda_1 \geq \ldots \geq \lambda_k = 0 \). The extrinsic mean \( \mu_E \) exists if and only if \( \lambda_m > \lambda_{m+1} \), and then \( \mu_E = \sigma(F') \) where \( F = (U_1, \ldots, U_m) \).

For a proof see Bhattacharya and Bhattacharya (2012), pp. 140, 141.

### 3.4. Projective Shape Space \( P\Sigma^k_m \)

First, recall that the real projective space \( \mathbb{R}P^m \) is the space of all lines through the origin in \( \mathbb{R}^{m+1} \). Its elements are \([p] = \{\lambda p : \lambda \in \mathbb{R}\} \setminus \{0\}\) for all \( p \in \mathbb{R}^{m+1} \setminus \{0\} \). It is also conveniently represented as the quotient \( S^m/G \) where \( G \) is the two-point group \([e, -e] \), \( e \) being the identity map and \(-ep = -p \) \((p \in S^m)\).

That is, a line through \( p \) is identified with \([p/|p|, -p/|p|] \) \((p \in \mathbb{R}^{m+1} \setminus \{0\})\). As a consequence, there is a unique Riemannian metric tensor on \( \mathbb{R}P^m = S^m/G \) such that \( p \rightarrow [p, -p] \) is a Riemannian submersion, with \((u, v)_{\mathbb{R}P^m} = u'v\) for all vectors \( u, v \) in \( T_p\mathbb{R}P^m \). The geodesic distance is given by \( d_p([p], [q]) = \arccos(|p'q'|) \in [0, \pi/2] \), and the cut locus of \([p] \) is \( Cut([p]) = \{[q] : \cos(|p'q'|) = \pi/2\} \), so that the injectivity radius of \( \mathbb{R}P^m \) is \( \pi/2 \). Its sectional curvature is constant +1 (as it is of \( S^m \)). The exponential map
of \( T_p \mathbb{R}^m \) (and its inverse on \( \mathbb{R}^m \setminus \text{Cut}([p]) \)) can be easily expressed in terms of those for the sphere \( S^m \). We will use \([\ ]\) for both representations.

The so-called Veronese-Whitney embedding of \( \mathbb{R}^m \) into \( S(m+1, \mathbb{R}) \) is given by

\[
J([p]) = pp^t, \quad (p = (p_1, \cdots, p_{m+1})') \in S^m.
\]

It is clearly \( O(m+1) \)-equivariant, with the group action on \( \mathbb{R}^m \) as: \( A[p] = [Ap] \) \( (A \in O(m+1)) \).

Turning to landmarks based projective shapes, assume \( k > m+2 \). A frame of \( \mathbb{R}^m \) is a set of \( m+2 \) ordered points \( ([p_1], \cdots, [p_{m+2}]) \) such that every subset of \( m+1 \) of these points spans \( \mathbb{R}^m \), i.e., every subset of \( m+1 \) points of \( \{p_1, \cdots, p_{m+2}\} \) spans \( \mathbb{R}^{m+1} \). The standard frame of \( \mathbb{R}^m \) is \( ([e_1], [e_2], \cdots, [e_{m+1}], [e_1 + e_2 + \cdots + e_{m+1}]) \), where \( e_i \in \mathbb{R}^{m+1} \) has 1 in the \( i \)th position and zeros elsewhere. A \( k \)-ad \( y = (y_1, \cdots, y_k) = ([p_1], \cdots, [p_k]) \in (\mathbb{R}^m)^k \) is in general position if there exist \( i_1 < i_2 < \cdots < i_{m+2} \) such that \( (y_{i_1}, \cdots, y_{i_{m+2}}) \) is a frame of \( \mathbb{R}^m \). A projective transformation \( \alpha \) on \( \mathbb{R}^m \) is defined by

\[
\alpha[p] = [Ap], \quad (p \in \mathbb{R}^{m+1} \setminus \{0\})
\]

where \( A \in GL(m+1, \mathbb{R}) \). The usual operation of matrix multiplication on \( GL(m+1, \mathbb{R}) \) then leads to a corresponding group of projective transformations on \( \mathbb{R}^m \). This is the projective group \( PGL(m) \). Note that, for a given \( A \in GL(m+1, \mathbb{R}) \), \( cA \) determines the same element of \( PGL(m) \) for all \( c \neq 0 \). The projective shape of a \( k \)-ad \( y = (y_1, \cdots, y_k) = ([p_1], \cdots, [p_k]) \in (\mathbb{R}^m)^k \) in general position is its orbit under \( PGL(m) \):

\[
\sigma(y) = \{ ay \equiv (\alpha[p_1], \cdots, \alpha[p_k]) : \alpha \in PGL(m) \},
\]

\[
(y = ([p_1], \cdots, [p_k] \text{ in general position}).
\]

The projective shape space \( PGS^k_m \) is the set of all projective shapes of \( k \)-ads in general position. Following Mardia and Patrangenaru (2005) and Patrangenaru et al. (2010), we will consider a particular dense open subset of \( PGS^k_m \). Fix a set of \( m+2 \) indices \( I = \{i_1 : j = 1, \cdots, m+2\}, 1 \leq i_1 < i_2 < \cdots < i_{m+2} \leq k \). Define \( PGI^k_m \) as the set of shapes \( \sigma(y) \) in \( PGS^k_m \), \( y = (y_1, \cdots, y_k) = ([p_1], \cdots, [p_k]) \), such that every subset of \( m+1 \) points of \( \{p_{i_j}\}, j = 1, \cdots, m+2 \) spans \( \mathbb{R}^m \).

The shape space \( PGI^k_m \) (with \( I = \{1, 2, \cdots, m+2\} \)) may be identified with \( (\mathbb{R}^m)^{k-m-2} \) (See Mardia and Patrangenaru (2005)).

4. Asymptotic Distribution Theory for Fréchet Means.

This section is devoted to the asymptotic distribution theory of sample Fréchet means, which lies at the heart of statistical inference based on Fréchet means. We first present a result which is broadly applicable to distributions on manifolds as well as more general locally Euclidean spaces such as stratified spaces. The basic idea behind it is rather simple. Suppose a probability \( Q \) on a metric space \( (S, \rho) \) has a Fréchet mean \( \mu \). Assume also the sample Fréchet mean \( \mu_n \) converges to it (a.s. or in probability), which is true in particular under the topological assumption (2). If, in local coordinates, \( \mu \) and \( \mu_n \) are expressed as \( \nu \) and \( \nu_n \) in an open subset of \( \mathbb{R}^e \) for some \( e \), then the Fréchet function \( F_0 \) of \( Q_n \), expressed in local coordinates as \( F_n \), say, satisfies a first order condition: \( \text{grad} F_n(\nu_n) = 0 \). A Taylor expansion of the left side around \( \nu \), one expresses \( \nu_n - \nu \) approximately as \( -\Delta^{-1}(\nu) \text{grad} F_n(\nu) \), where \( \Delta \) is the Hessian of \( \tilde{F} \) at \( \nu \). Since grad \( \tilde{F}_n(\nu) \) is the average of \( n \) \( \nu \)-dimensional i.i.d. random vectors, the classical CLT is applied to show that \( \sqrt{n}[\nu_n - \nu] \) is asymptotically normal. Here is the precise statement. For a detailed
proof see Bhattacharya and Lin (2017), Theorem 3.3. A slightly weaker version appears in Bhattacharya and Lin (2013).

Let \((S, \rho)\) be a metric space and \(Q\) a probability measure on its Borel \(\sigma\)-field. As before, define the Fréchet function of \(Q\) as

\[
F(p) = \int \rho^2(p, q) Q(dq) \quad (p \in S).
\]

Assume that \(F\) is finite on \(S\) and has a unique minimizer \(\mu = \arg\min_p F(p)\). Then \(\mu\) is called the Fréchet mean of \(Q\) (with respect to the distance \(\rho\)). Under broad conditions, the Fréchet sample mean \(\mu_n\) of the empirical distribution \(Q_n = \frac{1}{n} \sum_{j=1}^n \delta_{Y_j}\) based on independent \(S\)-valued random variables \(Y_j (j = 1, \ldots, n)\) with common distribution \(Q\) is a consistent estimator of \(\mu\). That is, \(\mu_n \to \mu\) almost surely, as \(n \to \infty\). Here \(\mu_n\) may be taken to be any measurable selection from the (random) set of minimizers of the Fréchet function of \(Q_n\), namely, \(F_n(p) = \frac{1}{n} \sum_{j=1}^n \rho^2(p, Y_j)\) (See Theorem 2.1).

The following assumptions are used in the proof of Theorem 4.2.

(A1) The Fréchet mean \(\mu\) of \(Q\) is unique.

(A2) \(\mu \in G\), where \(G\) is a measurable subset of \(S\), and there is a homeomorphism \(\phi : G \to U\), where \(U\) is an open subset of \(\mathbb{R}^s\) for some \(s \geq 1\) and \(G\) is given its relative topology on \(S\). The function

\[
x \mapsto h(x; q) := \rho^2(\phi^{-1}(x), q)
\]

is twice continuously differentiable on \(U\), for every \(q\) outside a \(Q\)-null set.

(A3) \(P(\mu_n \in G) \to 1\) as \(n \to \infty\).

(A4) Let \(D_r h(x; q) = \partial h(x; q) / \partial x_r, r = 1, \ldots, s\). Then

\[
E|D_r h(\phi(\mu); Y_1)|^2 < \infty, \quad E|D_{r,r'} h(\phi(\mu); Y_1)| < \infty \quad \text{for} \quad r, r' = 1, \ldots, s.
\]

(A5) Let \(u_{r,r'}(\epsilon; q) = \sup \{|D_{r,r'} h(\theta; q) - D_{r,r'} h(\phi(\mu); q)| : |\theta - \phi(\mu)| < \epsilon\}\). Then

\[
E|u_{r,r'}(\epsilon; Y_1)| \to 0 \quad \text{as} \quad \epsilon \to 0 \quad \text{for all} \quad 1 \leq r, r' \leq s.
\]

(A6) The matrix \(\Lambda = [ED_{r,r'} h(\phi(\mu); Y_1)]\) is nonsingular.

**Remark 4.1.** Observe that \(E h(x, Y_1) = F(\phi^{-1}(x)) = ED_r h(x, Y_1) = D_r F(\phi^{-1}(x))\), \(1 \leq r \leq s, x \in U\). Also, \(ED_r h(\phi(\mu), Y_1) = D_r F(\phi^{-1}(x))|_{x=\phi(\mu)} = 0, 1 \leq r \leq s\), since \(F(\phi^{-1}(x))\) attains a minimum at \(x = \phi(\mu)\).

**Theorem 4.2** (Bhattacharya and Lin, 2017). Under assumptions (A1)-(A6), \(f\)

\[
n^{1/2}[\phi(\mu_n) - \phi(\mu)] \xrightarrow{D} N(0, \Lambda^{-1} C \Lambda^{-1}), \quad \text{as} \quad n \to \infty,
\]

where \(C\) is the covariance matrix of \(\{D_r h(\phi(\mu); Y_1), r = 1, \ldots, s\}\).

**Proof.** The function \(x \to F_n(\phi^{-1}(x)) = \frac{1}{n} \sum_{j=1}^n h(x, Y_j)\) on \(U\) attains a minimum at \(\phi(\mu_n) \in U\) for all sufficiently large \(n\) (almost surely). For all such \(n\) one therefore has the first order condition

\[
\nabla F_n(\phi^{-1}(\mu_n)) = \frac{1}{n} \sum_{j=1}^n \nabla h(\mu_n, Y_j) = 0.
\]
where $\nu = \phi(\mu), \nu_n = \phi(\mu_n)$ (column vectors in $U$). Here $\nabla$ is the gradient $(D_1, \ldots, D_r)$. A Taylor expansion yields

$$0 = \frac{1}{n} \sum_{j=1}^{n} \nabla h(\nu_n, Y_j) = \frac{1}{n} \sum_{j=1}^{n} \nabla h(\nu, Y_j) + \Lambda_n(\nu_n - \nu)$$

where $\Lambda_n$ is the $s \times s$ matrix given by

$$\Lambda_n = \frac{1}{n} \sum_{j=1}^{n} [D_{r,r'} h(\theta_{n,r,r'}, Y_j)]_{r,r'=1,\ldots,s},$$

and $\theta_{n,r,r'}$ lies on the line segment joining $\nu_n$ and $\nu$. We will show that

$$\Lambda_n \to \Lambda \text{ in probability, as } n \to \infty.$$  

Fix $r, r' \in \{1, \ldots, s\}$. For $\delta > 0$, write $Eu_{r,r'}(\delta, Y_1) = \gamma(\delta)$. There exists $n = n(\delta)$ such that $P(|\nu_n - \nu| > \delta) < \delta$ for $n > n(\delta)$.) Now

$$E\left[\left| \frac{1}{n} \sum_{j=1}^{n} D_{r,r'} h(\nu_n, Y_j) - \frac{1}{n} \sum_{j=1}^{n} D_{r,r'} h(\nu, Y_j) \right| \cdot 1_{[|\nu_n - \nu| \leq \delta]} \right] \leq E \frac{1}{n} \sum_{j=1}^{n} |u_{r,r'}(\delta, Y_j)|$$

$$= Eu_{r,r'}(\delta, Y_1) = \gamma(\delta) \to 0$$

as $\delta \to 0$. Hence, by Chebyshev’s inequality for first moments, for $n > n(\delta)$ one has for every $\epsilon > 0$,

$$P\left( \left| \frac{1}{n} \sum_{j=1}^{n} D_{r,r'} h(\nu_n, Y_j) - \frac{1}{n} \sum_{j=1}^{n} D_{r,r'} h(\nu, Y_j) \right| > \epsilon \right) \leq \delta + \gamma(\delta)/\epsilon \to 0 \text{ as } \delta \to 0.$$  

This shows that

$$\frac{1}{n} \sum_{j=1}^{n} D_{r,r'} h(\nu_n, Y_j) - \frac{1}{n} \sum_{j=1}^{n} D_{r,r'} h(\nu, Y_j) \to 0 \text{ in probability as } n \to \infty.$$  

Next, by the strong law of large numbers,

$$\frac{1}{n} \sum_{j=1}^{n} D_{r,r'} h(\nu, Y_j) \to ED_{r,r'} h(\nu, Y_1) \text{ almost surely, as } n \to \infty.$$  

Since (32) – (34) hold for all $r, r'$, (31) follows. The set of symmetric $s \times s$ positive definite matrices is open in the set of all $s \times s$ symmetric matrices, so that (31) implies that $\Lambda_n$ is nonsingular with probability going to 1 and $\Lambda_n^{-1} \to \Lambda^{-1}$ in probability, as $n \to \infty$. Note that $E\nabla h(\nu, Y_1) = 0$ (see Remark 4.1). Therefore, using (A4), by the classical CLT and Slutsky’s Lemma, (29) leads to

$$\sqrt{n}(\nu_n - \nu) = \Lambda_n^{-1} - (1/\sqrt{n}) \frac{1}{n} \sum_{j=1}^{n} \nabla h(\nu, Y_j) \xrightarrow{D} N(0, \Lambda^{-1} C A^{-1}),$$

as $n \to \infty$. \hfill \qedsymbol

For the case of the extrinsic mean, let $M$ be a $d$-dimensional differentiable manifold, and $J : M \to E^N$ an embedding of $M$ into an $N$-dimensional Euclidean space. Assume that $J(M)$ is closed in $E^N$, which is always the case, in particular, if $M$ is compact. The extrinsic distance $\rho_{E,J}$ on $M$ is defined as $\rho_{E,J}(p, q) = |J(p) - J(q)|$ for $p, q \in M$, where $| \cdot |$ denotes the Euclidean norm of $E^N$. The image $\mu$ in $J(M)$ of the extrinsic mean $\mu_{E,J}$ is then given by $\mu = P(m)$, where $m$ is the usual mean of $Q \circ J^{-1}$ thought of as a probability on the Euclidean space $E^N$, and $P$ is the orthogonal projection defined on
an $N$-dimensional neighborhood $V$ of $m$ into $J(M)$ minimizing the Euclidean distance between $\mu \in V$ and $J(M)$. If the projection $P$ is unique on $V$ then the projection $\mu_n = P(m_n)$ of the Euclidean mean $m_n = \sum_{j=1}^{n} J(Y_j)/n$ on $J(M)$ is, with probability tending to one as $n \to \infty$, unique and lies in an open neighborhood $G$ of $\mu = P(m)$ in $J(M)$. Theorem 4.2 immediately implies the following result of Bhattacharya and Patrangenaru (2003) (Also see Bhattacharya and Bhattacharya (2012), Proposition 4.3). Assume that $P$ is uniquely defined in a neighborhood of the $N$-dimensional Euclidean mean $m$ of $Q \circ J^{-1}$.

Let $\phi$ be a diffeomorphism on a neighborhood $G$ of $\mu = P(m)$ in $J(M)$ onto an open set $U$ in $\mathbb{R}^d$. Then, using the notation of (27),

$$\sqrt{n} [\phi(\mu_n) - \phi(\mu)] = \sqrt{n} [\phi(P(m_n)) - \phi(P(m))] \overset{d}{\to} N(0, \Lambda^{-1} C \Lambda^{-1})$$

as $n \to \infty$. One may, in particular, choose $(U, \phi)$ to be a coordinate neighborhood of $\mu = P(m)$ in $J(M)$. In Bhattacharya and Patrangenaru (2003), however, $\phi$ is chosen to be the linear orthogonal projection on $G$ into the tangent space $T_\mu J(M)$.

For a more computable expression of the limit, let $X_j, 1 \leq j \leq n$, be i.i.d, $M$-valued observations with common distribution $Q$, and $Y_j = J(X_j), 1 \leq j \leq n$. In a neighborhood $V$ of $m$, the differential $d_j P$ maps $T_j E^N \approx E^N$. One expresses $d_{P(m)} e_j = \sum_{i=1}^{d} b_{ij} F_i$,

$$d_m (\bar{Y} - m) = \sum_{j=1}^{d} \sum_{i=1}^{N} b_{ij} (\bar{Y} - m)(j) F_i = \sum_{i=1}^{d} \left( \sum_{j=1}^{N} b_{ij} (\bar{Y} - m)(j) \right) F_i \quad (F_i = F_i(P(m))).$$

Thus one arrives at the following result.

**Proposition 4.3.** Assume the projection $P$ is uniquely defined and is continuously differentiable in a neighborhood $V$ of $m = EY_j$, and $E|Y_j|^2 < \infty$. Then

(36)$$\sqrt{n} d_m P(\bar{Y} - m) \overset{d}{\to} N(0, \Sigma)$$

where $\Sigma = B'C B$ with $b = ((b_{ij}))$ and $C$ is the $N \times N$ covariance matrix of $Y_j$.

**Corollary 4.4** (CLT for Intrinsic Means-I). Let $(M, g)$ be a $d$-dimensional complete Riemannian manifold with metric tensor $g$ and geodesic distance $\rho_g$. Suppose $Q$ is a probability measure on $M$ with intrinsic mean $\mu_1$, and that $Q$ assigns zero mass to a neighborhood, however small, of the cut locus of $\mu_1$. Let $\phi = \text{Exp}_{\mu_1}^{-1}$ be the inverse exponential, or log-, function at $\mu_1$ defined on a neighborhood $G$ of $\mu = \mu_1$ onto its image $U$ in the tangent space $T_{\mu_1}(M)$. Assume that the assumptions (A4)-(A6) hold. Then, with $s = d$, the CLT (27) holds for the intrinsic sample mean $\mu_n = \mu_{n,1}$, say.

**Remark 4.5.** In addition to providing a CLT for manifolds (of dimension $d$), Theorem 4.2 applies to many stratified spaces which are manifolds of different dimensions $s$ glued together. See Bhattacharya and Lin (2017) for a simple derivation of a CLT for the so-called Open Book model, originally due to Hotz et al. (2013). Another stratified space to which Theorem 4.2 applies is $\Sigma_m^{k}, m > 2, k > m$, described in Section 3 (see Remark 4.17).

**Remark 4.6.** For manifolds of dimension $d$ (i.e., $s = d$), Theorem 4.2 and Corollary 4.4 improve upon Theorem 2.3 and 5.3 in Bhattacharya and Bhattacharya (2012) (and earlier results in Bhattacharya and Patrangenaru (2005)).
We now turn to the derivation of the asymptotic distribution of sample intrinsic Fréchet means on Riemannian manifolds which does not require the support restriction of Corollary 4.4.

For the case of the circle $S^1$, necessary and sufficient conditions for the existence of the intrinsic mean was established in Bhattacharya (2007), under the assumption of a continuous density with respect to the uniform distribution. The result also along with a central limit theorem for the sample intrinsic mean in Bhattacharya and Bhattacharya (2012), pp. 31-34, 72-75. A proof of the CLT was also obtained independently in McKilliam et al. (2012). Some additional results, especially for distributions with discontinuous density may be found in the recent article Hotz and Huckemann (2015).

**Proposition 4.7.** Suppose a complete orientable $d$-dimensional Riemannian manifold $(M, g)$ has the property that the image $D \subset T_pM$ of $M \setminus \text{Cut}(p)$ under the map $\log_p = \text{Exp}_p^{-1}$ is the same for all $p \in M$, and the push forward of the volume measure on $D$ under the $\log_p$ map is also the same for all $p$. Assume that the intrinsic mean $\mu_i$ of a probability $Q$ on $M$ exists, and that $Q$ is absolutely continuous in a neighborhood $W$ of $\text{Cut}(\mu_1)$ with a density $f$ on $W$ which is twice continuously differentiable. Assume also that the first and second derivatives of $p \to f(\text{Exp}_p v)$, in local coordinates, are bounded for $p$ in a neighborhood of $\mu_1$ by functions $f_i(v), f_{ij}(v)$ such that, for a sufficiently small $\epsilon > 0$,

$$\int_{\{R-\epsilon<|v|<R\}} |v|^2 f_i(v)m(dv) < \infty, \quad \int_{\{R-\epsilon<|v|<R\}} |v|^2 f_{ij}(v)m(dv) < \infty, \quad (i, j = 1, \ldots, d),$$

where $m(dv)$ is the push forward on $T_pM$ of the volume measure by the map $\log_p \mu_1$. Then there exists a neighborhood of $\mu_1$ in which the Fréchet function (1) with $\rho = \rho_2$, is twice continuously differentiable.

**Proof.** First note that there exist $r > 0$ and $\epsilon > 0$, both sufficiently small and a geodesic ball $B_r$ with center $\mu_1$ and radius $r > 0$ such that $(\text{Cut}(B_r))^c \subset W$, where $\text{Cut}(B_r) = \cup \{ \text{Cut}(p) : p \in B_r \}$ and $A^\epsilon$ is the $\epsilon$-neighborhood of a set $A \subset M$. For $p \in B_r$,

$$F(p) = \int_{\{|q|:|\log_p q|<R-\epsilon\}} \rho_2^2(p, q)Q(dq) + \int_{\{|q|:|\log_p q|<R-\epsilon\}} \rho_2^2(p, q)Q(dq)$$

The first integral in (38) is clearly twice continuously differentiable. The second integral may be expressed as

$$\int_{\{R-\epsilon<|v|<R\}} |v|^2 f(\text{Exp}_p v)m(dv),$$

where $R = \rho_2(p, \text{Cut}(p))$.

**Remark 4.8.** We conjecture that the conclusion of Proposition 4.7 holds much more generally and, in particular, for all compact orientable Riemannian manifolds, if $Q$ has a twice continuously differentiable density.

The following result is an immediate consequence of Proposition 4.7.

**Corollary 4.9.** Let $M = S^d$ with the usual Riemannian metric tensor, and $Q$ a probability measure on it. (a) Then the Fréchet function is twice continuously differentiable if $Q$ has a twice continuously differentiable density. (b) If $Q$ is absolutely continuous in a neighborhood of the cut locus $\text{Cut}(p)$ of a point $p$, with a twice continuously differentiable density
there, then the Fréchet function is twice continuously differentiable in a neighborhood of \( p \).

For the statement of the next result we continue to use the notation of Corollary 4.4.

**Theorem 4.10** ((Bhattacharya and Lin, 2017)). Suppose that the intrinsic mean \( \mu \) of \( Q \) exists, and that \( Q \) is absolutely continuous in a neighborhood \( W \) of the cut locus of \( \mu \) with a continuous density with respect to the volume measure. Assume also that (i) \( Q(\text{Cut}(B(\mu; \epsilon))) = O(\epsilon^{d-c}) \), \( \epsilon \to 0 \), for some \( c, 0 \leq c < d \), (ii) on some neighborhood \( V \) of \( \nu = \phi(\mu) = 0 \) the function \( \theta \to F(\phi^{-1}(\theta)) \) is twice continuously differentiable with a nonsingular Hessian \( \Lambda(\theta) \), and (iii) (A4) holds with \( \phi(\mu) \) replaced by \( \theta, \forall \theta \in V \). Then, if \( d > c + 2 \), one has the CLT (27) for the sample intrinsic mean \( \mu_n \).

**Proof.** One may take the neighborhood \( V \) of \( \nu = 0 \) sufficiently small such that \( \text{Cut}(\phi^{-1}(V)) \subset W \). Then \( Z_n(\theta) := n^{-1} \sum_{1 \leq j \leq n} \text{grad} \ h(\theta, Y_j) \) is well defined for \( Y_j \notin \text{Cut}(\phi^{-1}\theta) \), \( j = 1, \ldots, n \), that is, with probability one, provided \( \theta \in V \), since \( Q(\text{Cut}(\phi^{-1}\theta)) = 0 \). By the classical CLT, \( Z_n(0) := n^{-1} \sum_{1 \leq j \leq n} \text{grad} \ h(0, Y_j) \) is of the order \( O_p(n^{-1/2}) \). Let \( B_n \) be the ball in \( T_p \) with center \( \nu = \phi(\mu) = 0 \) and radius \( n^{-1/2} \log n \). By hypothesis, the probability that \( Y_j \notin \text{Cut}(\phi^{-1}(B_n)) \) is \( O((n^{-1/2} \log n)^{d-c}) \). For \( \phi^{-1}(B_n) \) is the geodesic ball \( B(\mu; n^{-1/2} \log n) \), hence the probability that the set \( \{ Y_j : j = 1, \ldots, n \} \) intersects \( \text{Cut}(\phi^{-1}(B_n)) \) is \( O(n(n^{-1/2} \log n)^{d-c} = o(1) \) if \( d > c + 2 \). Therefore, with probability converging to 1, one may use a Taylor expansion of \( Z_n(\theta) \) in \( B_n \),

\[
(40) \quad Z_n(\theta) = Z_n(\nu) + \Lambda_n(\theta)(\theta - \nu), \quad (\theta \in B_n), \quad (\nu = 0),
\]

where \( \Lambda_n(\theta) \) is the \( d \times d \) matrix whose \((r, r')\) element is \( n^{-1} \sum_{1 \leq j \leq n} D_{rr'} h(\theta; Y_j, Y_j) \) lying on the line segment joining \( \theta \) and \( \nu = 0 \). By hypothesis (ii), with probability converging to one as \( n \to \infty \), \( \Lambda_n(\theta) \) is nonsingular for all large \( n \) \( (\theta \in B_n) \) since its difference (in norm) from the Hessian \( \Lambda(\theta) \) goes to zero as \( n \to \infty \), by the strong law of large numbers. Also, with probability going to 1, the function \( \theta \to H_n(\theta) = 0 - \Lambda_n(\theta)^{-1} Z_n(\nu) \) maps \( B_n \) into itself, where \( B_n \) is the closure of \( B_n \). For this argument recall that \( Z_n(0) = O_p(n^{-1/2}) \) by the classical CLT. By the Brouwer fixed point theorem (Milnor, 1965), \( H_n(\theta) \) has a fixed point. Let \( \nu_n \) denote a measurable selection from the set of fixed points in \( \bar{B}_n \). It follows that, with probability going to 1, \( \nu_n \) converges to \( \nu \) and satisfies the first order equation (28), and \( \nu_n \) is the sample intrinsic mean, since the Fréchet function is strictly convex in a neighborhood of \( \nu \). The CLT now follows as in the last sentence and relation (35) of the proof of Theorem 4.2.

\[
\square
\]

**Corollary 4.11.** Suppose \( Q \) on \( M = S^d \) \((d > 2)\) has an intrinsic mean \( \mu \) and is absolutely continuous on a neighborhood \( W \) of \( \text{Cut}(\mu) \) with a continuous density on \( W \). Then the CLT for the sample intrinsic mean holds.

**Remark 4.12.** It follows from the proof of Theorem 2.2, specialized to \( S^d \), that the Hessian \( \Lambda \) is positive definite.

**Remark 4.13.** Although it is curious that the proof of Theorem 4.10 does not hold for \( d = 2 \), the authors expect that a proof of Corollary 4.11 for the case \( d = 2 \) may be given using polar coordinates. For the moment, the CLT for \( S^2 \) is derived only under the support restriction of Corollary 4.4.

**Remark 4.14.** Suppose \( G \) is a Lie group of isometries on \( S^d \), \( d > 2 \). Then the projection \( \pi : S^d \to S^d / G \) is a Riemannian submersion on \( S^d \) onto its quotient space \( M = S^d / G \).
Let $\hat{Q}$ be a probability measure on $S^d$ with a twice continuously differentiable density and a Karcher or intrinsic mean $\mu$. Let $\bar{\mu}$ be the projection of $\mu$ and $\bar{\mu}$ that of the sample intrinsic (or Karcher) mean $\mu_n$. Then, in local coordinates, the differential of the Fréchet function on $M$ vanishes at $\bar{\mu}$, because $\pi$ is smooth and the differential of the Fréchet function on $S^d$ vanishes at $\mu$. The delta method provides a CLT for the corresponding sample Fréchet mean $\bar{\mu}_n$ in local coordinates. However, $\bar{\mu}$ $(\bar{\mu}_n)$ are unlikely to be the intrinsic (respectively, sample intrinsic) mean of $\hat{Q}$ (respectively, $\hat{Q}_n$) by the projection map.

**Remark 4.15.** One may also explore the opposite route for a probability $\hat{Q}$ on $M$ with a density and a unique intrinsic/Karcher mean $\hat{\mu}$ and a probability $\rho$, among a fairly large family of distributions with smooth densities on $S^d$ whose projection on $M$ is $\hat{Q}$, such that $\hat{Q}$ satisfies the hypothesis of Corollary 4.11 with $\pi(\mu) = \hat{\mu}$. One may then apply the CLT on $S^d$ to derive one on $S^d/\mathcal{G}$. As an example consider the antipodal map $q(p) = -p$, and $\mathcal{G} = \{g, \text{identity}\}$. Let $\hat{Q}$ be a probability on $M = S^d/\mathcal{G} = \mathbb{R}P^d$ (the real projective space) thought of as a probability on the upper hemisphere vanishing smoothly at the boundary, and with a unique intrinsic mean $\bar{\mu} = \{\mu, -\mu\}$, where $\mu$ is the Karcher mean of $\hat{Q}$ (restricted to the hemisphere). This opens a way for CLT’s on Kendall’s shape spaces as well.

**Remark 4.16.** As indicated in Remark 4.14, one of the significances of a CLT on $S^d$ is that it may provide a route to intrinsic CLTs on $S^d/\mathcal{G}$, the space of orbits under a Lie group $\mathcal{G}$ of isometries of $S^d$. Such spaces include the so-called axial spaces (or real projective spaces $\mathbb{R}P^d$), and Kendall type shape spaces which are important in shape-based image analysis. For the latter spaces, $S^d$ is the so-called preshape sphere. Observe that the hypothesis (i) of Theorem 4.10 may not hold in all such quotient spaces. For example, on $\mathbb{R}P^d$ one only has the order $O(\epsilon)$ in hypothesis (i) in Theorem 4.10, since the cut locus of the a point in $\mathbb{R}P^d$ is isomorphic to $\mathbb{R}P^{d-1}$. For Kendall’s planar shape space, identified as the complex projective space $\mathbb{C}P^{k-2}$, of dimension $d = 2k - 4$, the volume measure of Cut $(B(\mu; \epsilon))$ is $O(\epsilon^2)$, since the cut locus of a point of $\mathbb{C}P^{k-2}$ is isomorphic to $\mathbb{C}P^{k-3}$. For these facts one may refer to Gallot et al. (1990), Section 2.114, pp. 102, 103.

**Remark 4.17.** For $m > 2$, $k > m$, $\Sigma_m^k$ is a stratified space in the intrinsic topology. But the projection $S^{m(k-1)-1} \to \Sigma_m^k$ (see (6)) is continuous and $\Sigma_m^k$ is a compact metric space. Hence Theorem 4.2 still applies to this stratified space.

5. **Nonparametric Inference on General Manifolds**

Theorems 2.1, 4.10 allow us to construct nonparametric confidence regions for intrinsic and extrinsic means of probability measures $\hat{Q}$ on a manifold $M$, and to carry out nonparametric two-sample tests for the equality of such means of two distributions $\hat{Q}_1$ and $\hat{Q}_2$ on $M$. The latter tests are really meant to distinguish $\hat{Q}_1$ from $\hat{Q}_2$. On high dimensional spaces, such as the shape spaces of main interest here, the means are generally good indices for this purpose, as the data examples in Section 10 show.

For the construction of an extrinsic confidence region for the extrinsic mean $\mu_F$ of $\hat{Q}$ one may use the corresponding region for $\mu_J$ using (36) and then transform by $J^{-1}$. The following asymptotic chisquare distribution is an easy consequence of Proposition 4.3:

\[
\chi^2 \quad (d \Sigma P)(\bar{Y} - m) \quad (\bar{B} \Sigma \bar{B})^{-1} [(d \Sigma P)(\bar{Y} - m)] \to \chi^2_d \quad \text{in distribution},
\]

where $\chi^2_d$ is the chisquare distribution with $d$ degrees of freedom. Here $\bar{B} = B(\bar{Y})$ estimates $B = B(m)$, and $\Sigma$ is the sample covariance matrix of $Y_1, \cdots, Y_n$. The statistic does
not depend on the choice of the orthonormal basis of $T_\gamma(J(M))$ for computing $\hat{B}$. The relation (41) may be used to construct extrinsic mean $\mu_E = J^{-1}P(m)$. Bootstrapping, which leads to a smaller order of coverage error in the case of an absolutely continuous $Q$, may not always be feasible if $N$ is large and the sample size $n$ is not sufficiently large to ensure that, with high probability, the bootstrap estimate of the sample covariance matrix is not singular.

Turning to the (local) intrinsic mean $\mu_I$ of $Q$, Theorem 4.10 leads to the asymptotic chisquare distribution

$$n[\phi(\mu_n) - \phi(\mu_I)]^T \hat{\Lambda}\hat{\Sigma}^{-1}\hat{\Lambda}[\phi(\mu_n) - \phi(\mu_I)] \to \chi^2(d)$$

in distribution as $n \to \infty$, where $\phi$ denotes an estimate with $Q$ replaced by the empirical $Q_n$; that is, the distribution $Q \circ \phi^{-1}$ of $Y_1$ is replaced by $Q_n \circ \phi^{-1} = n^{-1}\sum_{1 \leq i \leq n} \delta_{Y_i}$. This leads to a confidence region for $\mu_I$.

We next consider the the two-sample problem of distinguishing two distributions $Q_1$ and $Q_2$ on $M$, based on two independent samples of sizes $n_1$ and $n_2$, respectively: $\{Y_{j_1} = J(X_{j_1}) : j = 1, \cdots, n_1\}, \{Y_{j_2} = J(X_{j_2}) : j = 1, \cdots, n_2\}$. Hence the proper null hypothesis is $H_0 : Q_1 = Q_2$. For high dimensional $M$ it is often sufficient to test if the two Fréchet means are equal. For the extrinsic procedure, again consider an embedding $J$ into $E^N$. Write $\mu_i$ for $\mu_i^E$ for the population means and $\bar{Y}_i$ for the corresponding sample means on $E^N (i = 1, 2)$. Let $n = n_1 + n_2$, and assume $n_1/n \to p_1, n_2/n \to p_2 = 1 - p_1$, $0 < p_i < 1 (i = 1, 2)$, as $n \to \infty$. If $\mu_1 \neq \mu_2$ then $Q_1 \neq Q_2$. One may then test

$$H_0 : \mu_1 = \mu_2 \text{ based on } \bar{Y}_1 - \bar{Y}_2 \text{ is generally not a good test.}$$

Instead, we compare the two extrinsic means $\mu_i^E$ and $\mu_E$ of $Q_1$ and $Q_2$ and test for their equality. This is equivalent to testing if $P(\mu_1) = P(\mu_2)$. Then, by (41), assuming $H_0$,

$$n^{1/2}d_Y P(\bar{Y}_1 - \bar{Y}_2) \to N(0, B(p_1\Sigma_1 + p_2\Sigma_2)B')$$

in distribution, as $n \to \infty$. Here $\bar{Y} = p_1\bar{Y}_1 + p_2\bar{Y}_2$ is the pooled estimate of the common mean $\mu_1 = \mu_2 = \mu$, say, $B = B(\mu)$ (see Proposition 4.3.), and $\Sigma_1, \Sigma_2$ are the covariance matrices of $\bar{Y}_1$ and $\bar{Y}_2$. This leads to the asymptotic chisquare statistic below:

$$n[d_Y P(\bar{Y}_1 - \bar{Y}_2)]^T [B(p_1\hat{\Sigma}_1 + p_2\hat{\Sigma}_2)B']^{-1} [d_Y P(\bar{Y}_1 - \bar{Y}_2)] \to \chi^2_d$$

in distribution, as $n \to \infty$. Here $\hat{B} = B(\bar{Y}), \hat{\Sigma}_i$ is the sample covariance matrix of $Y_{j_1}$. One rejects the null hypothesis $H_0$ at a level of significance $1 - \alpha$ if and only if the observed value of the left side of (44) exceeds $\chi^2_d(1 - \alpha)$.

For the two-sample intrinsic test, let $\mu_{i_1}, \mu_{i_2}$ denote the intrinsic means of $Q_1$ and $Q_2$ and consider $H_0 : \mu_{i_1} = \mu_{i_2}$. Denoting by $\mu_{i_1}, \mu_{i_2}$ the intrinsic sample means, (44) implies that, under $H_0$,

$$n^{1/2}[\phi(\mu_{i_1}) - \phi(\mu_{i_2})] \to N(0, p_1\Lambda_{i_1}^{-1}\hat{\Sigma}_i\hat{\Lambda}_{i_1}^{-1} + p_2\Lambda_{i_2}^{-1}\hat{\Sigma}_2\hat{\Lambda}_{i_2}^{-1})$$

in distribution, where $\phi = Exp_p^{-1}$ for some convenient $p$ in $M$, and $\Lambda_i, \hat{\Sigma}_i$ are as in Theorem 4.10 and (44) with the empirical $Q_n$ in place of $Q_i (i = 1, 2)$. One simple choice for $p$ is the pooled estimate $\mu_n = p_1\mu_{i_1} + p_2\mu_{i_2}$, and with this choice we write $\phi$ for $\phi$. The test then rejects $H_0 : Q_1 = Q_2$, if

$$n[\phi(\mu_{i_1}) - \phi(\mu_{i_2})]^T [p_1\Lambda_{i_1}^{-1}\hat{\Sigma}_i\hat{\Lambda}_{i_1}^{-1} + p_2\Lambda_{i_2}^{-1}\hat{\Sigma}_2\hat{\Lambda}_{i_2}^{-1}]^{-1} [\phi(\mu_{i_1}) - \phi(\mu_{i_2})] \chi^2_d(1 - \alpha).$$

Finally, consider a match pair problem with i.i.d. observations $(X_{j_1}, X_{j_2})$ having the distribution $Q$ on the product manifold $M \times M$. If $J$ is an embedding of $M$ into $E^N$, then
\( \hat{J}(x, y) = (J(x), J(y)) \) is an embedding of \( M \times M \) into \( E^N \times E^N \). Let \( \mu_{E_1}, \mu_{E_2} \) be the extrinsic means of the (marginal) distributions \( Q_1 \) and \( Q_2 \) of \( X_j \) and \( X_{j_2} \), respectively. Once again, we are interested in testing \( H_0: Q_1 = Q_2 \) by checking if \( \mu_{E_1} = \mu_{E_2} \). Note that the extrinsic mean of \( Q \) is \( \tilde{\mu}_E = (\mu_{E_1}, \mu_{E_2}) \). If \( \bar{Y}_1, \bar{Y}_2 \) are the sample means of \( Y_{j_1} = J(X_{j_1}), Y_{j_2} = J(X_{j_2}), j = 1, \cdots, n \), on \( E^N \) with \( E(Y_{j_1}) = \mu_1 \) and \( E(Y_{j_2}) = \mu_2 \), and \( \bar{Y} = (\bar{Y}_1, \bar{Y}_2) \), then the extrinsic sample mean in the image space \( \hat{J}(M \times M) \) is \( (P(\bar{Y}_1), P(\bar{Y}_2)) \). Also, write \( \tilde{Y} = (\bar{Y}_1 + \bar{Y}_2)/2 \). Under \( H_0, \mu_1 = \mu_2 = \mu \), say, and one has

\[ n^{1/2}d_YP(\bar{Y}_1 - \bar{Y}_2) \rightarrow N(0, \Sigma_{11} + \Sigma_{22} - \Sigma_{12} - \Sigma_{21}). \]

On the right, \( \Sigma_{11} \) and \( \Sigma_{22} \) are the \( d \times d \) covariance matrices of \( (d_\mu P)(Y_{j_1} - \mu_1) \) and \( (d_\mu P)(Y_{j_2} - \mu_2) \), while \( \Sigma_{12} \) is the \( d \times d \) cross covariance matrix of \( (d_\mu P)(Y_{j_1} - \mu_1) \) and \( (d_\mu P)(Y_{j_2} - \mu_2) \), and \( \Sigma_{21} = \Sigma_{12}^T \). As above, one derives a chisquare test for \( H_0 \), using (47) and sample estimates of the covariance matrices.

### 6. Intrinsic and Extrinsic Analysis and Curvature

In this section we provide explicit expressions of asymptotic dispersions for the intrinsic CLT on general Riemannian manifolds, and relate this to curvature, with applications to the sphere and planar shape spaces.

For intrinsic analysis, consider the function \( h(z, y) = \rho_p^2(Exp_p z, Exp_p y) \) for \( z, y \) in \( T_p M \), with an appropriate choice of \( p \). One first needs to express explicitly the quantities \( D_h(z, y), D_s D_t h(z, y) \) in normal coordinates at \( p \), i.e., \( z = 0 = Exp_p^{-1} p \). For this let \( \gamma(s) \) be a (constant speed) geodesic starting at \( p \), and \( m \in M \). Define the parametric surface \( c(s, t) = Exp_m(t Exp_{m}^{-1} \gamma(s)), s \in [0, \epsilon), \epsilon > 0 \) small. Note that \( c(0, 0) = m \) for all \( s, c(0, 1) = \gamma(s) \), and that, for all fixed \( s \in [0, \epsilon), t \rightarrow c(s, t) \) is a geodesic starting at \( m \) and reaching \( \gamma(s) \) at \( t = 1 \). Writing \( T(s, t) = (\partial/\partial t) c(s, t), S(s, t) = (\partial/\partial s) c(s, t), \) one then has \( S(s, 0) = 0, S(s, 1) = \dot{\gamma}(s) \). Also, \( (T(s, t), T(s, t)) \) does not depend on \( t \) and, therefore,

\[ \rho_p^2(\gamma(s), m) = \int_0^1 \langle T(s, t), T(s, t) \rangle dt. \]

Differentiating this with respect to \( s \) and recalling the symmetry \( (D/\partial s)T(s, t) = (D/\partial t)S(s, t) \) on a parametric surface (See Do Carmo (1992), p. 68, Lemma 3.4), and \( (D/\partial t)T(s, t) = 0 \), one has

\[ (d/\partial s)\rho_p^2(\gamma(s), m) = 2 \int_0^1 \langle (D/\partial s)T(s, t), T(s, t) \rangle dt \]

\[ = 2 \int_0^1 \langle (D/\partial t)S(s, t), T(s, t) \rangle dt = 2 \int_0^1 (d/\partial t) \langle S(s, t), T(s, t) \rangle dt \]

\[ = 2\langle S(s, 1), T(s, 1) \rangle = -2\langle \dot{\gamma}(s), Exp_p^{-1} \gamma(s), m \rangle. \]

Setting \( s = 0 \) in (49) and letting \( \dot{\gamma}(0) = v_r \), with \( \{v_r : r = 1, \cdots, d\} \) an orthonormal basis of \( T_p M \), one shows that the normal coordinates \( y_r \) of \( m \) (i.e., the coordinates of \( y = Exp_p^{-1} m \) with respect to \( \{v_r : r = 1, \cdots, d\} \) satisfy

\[ -2y_r^r = -2\langle Exp_p^{-1} m, v_r \rangle = [(d/\partial s)\rho_p^2(\gamma(s), m)]_{s=0}. \]

From this one gets

\[ D_r h(0, y) = -2y_r^r (r = 1, \cdots, d). \]
If $Q(Cut(p)) = 0$, then writing $\tilde{Q}$ for the distribution induced from $Q$ by the map $Exp_p^{-1}$ on $T_p M$, the Fréchet function and its gradient in local coordinates may be expressed as

$$F(q) = \int p^2(y, m)Q(dm) = \int h(z, y)Q(dy) = \hat{F}(z),$$

where $z = Exp_p^{-1} q$ and $D_r \hat{F}(z) = -2 \int y' \tilde{Q}(dy)$. Since a (local) minimum of this is attained at $q = \mu_1$, $\hat{F}$ must satisfy a first order condition $D_r \hat{F}(z) = 0$ at $z = \nu$. In particular, letting $p = \mu_1$ and, consequently, $\nu = 0$, one has $\int D_r h(0, y)\tilde{Q}(dy) = 0$, so that (51) yields

$$\int y' \tilde{Q}(dy) = 0 \quad (r = 1, \ldots, d), \quad (\tilde{Q} = Q \circ \phi^{-1}, \phi = Exp_{\mu_1}^{-1}).$$

By Theorem 4.10, the asymptotic distribution of the sample intrinsic mean $\mu_n$ is that of $\phi^{-1}(\nu_n)$, where $\phi = Exp_{\mu_1}^{-1}$, and

$$\sqrt{n}(\nu_n - \nu) \sim \Lambda^{-1}(1/\sqrt{n}) \sum_{1 \leq j \leq n} D h(\nu, Y_j), \quad (\Lambda_{rs} = ED_r D_s h(\nu, Y_1), 1 \leq r, s \leq d),$$

with $Y_j = \phi(X_j)$, where $X_j$ are i.i.d. with distribution $Q$. By (51), the right side of (54) simplifies to $\Lambda^{-1}[-2(1/\sqrt{n}) \sum_{1 \leq j \leq n} Y_j]$, if $p = \mu_1$ (and $\nu = 0$).

For an analytical study of the Hessian $\Lambda$ of the Fréchet function, one derives from (63) the relation

$$\frac{d^2}{ds^2} \rho^2_2(\gamma(s), m) = 2\langle D_s T(s, 1), S(s, 1) \rangle = 2\langle D_t S(s, 1), S(s, 1) \rangle,$$

and let $C$ denote the supremum of all sectional curvatures of $M$ and let

$$f(t) = \begin{cases} 1 & \text{if } C = 0 \\ (\sqrt{C}t) \cos(\sqrt{C}t) / \sin(\sqrt{C}t) & \text{if } C > 0, \\ \sqrt{-Ct} \cosh(\sqrt{-Ct}) / \sinh(\sqrt{-Ct}) & \text{if } C < 0. \end{cases}$$

Also let $t_0$ be the supremum of all $t$ such that $f(t) > 0$. For $d \times d$ symmetric matrices $A$, $B$, the order relation $A \geq B$ means $A - B$ is nonnegative definite.

**Theorem 6.1** (Bhattacharya and Bhattacharya (2008)). Assume $|Y_1| = |\log_{\mu_1} X_1| \leq t_0$ a.s. In addition, if the hypotheses for the CLT in corollary 4.4 or Theorem 4.10 hold, one has

$$\Lambda = \langle (\Lambda_{ij}) \rangle \geq \left( 2E \left( \frac{1 - f|Y_1|}{|Y_1|^2} Y_1^i Y_1^j + f(|Y_1|)\delta_{ij} \right) \right),$$

with equality if the sectional curvature is consistent.

**Remark 6.2.** It is simple to check that on $M = S^d$, the Hessian $\Lambda$ given by the right side of (58), with $C = 1$, is nonsingular.

**Remark 6.3.** Kendall and Le (2011) obtained the exact expression for the Hessian for the intrinsic Fréchet mean on the important case of the planar shape space $\Sigma^k$, which has a constant holomorphic curvature.

**Remark 6.4.** Note that the relations in (50) provide the gradient of the intrinsic Fréchet function $F(p)$ in normal coordinates around $p$. 
Example 6.5. (Confidence region for the intrinsic mean of $Q$ on the sphere $S^d$). Let $\mu_I$ be the intrinsic mean of $Q$ on $S^d$. Given $n$ i.i.d. observations $X_1, \ldots, X_n$ on $S^d$ with common distribution $Q$, let $\mu_n$ be the intrinsic sample mean. Write $\phi = Exp_{\mu_I}^{-1}$, and $\phi_p = Exp_{p}^{-1}$, so that $\phi_{\mu_I} = \phi$. By Theorem 4.2,
\begin{equation}
\sqrt{n}(\phi(\mu_n) - \phi(\mu_I)) = \sqrt{n}(\phi(\mu_n)) \to N(0, \Lambda^{-1}\Sigma\Lambda^{-1}) \text{ in distribution as } n \to \infty,
\end{equation}
where the $d \times d$ matrices $\Sigma$ and $\Sigma$ are given by
\begin{equation}
\Sigma = 4\text{Cov}(\phi(X_1)),
\end{equation}
\begin{equation}
\Lambda_{rs} = 2E[(1 - (X^I_{rs})^2)^{-1} \{1 - (1 - (X^I_{rs})^2)^{-1/2} \cdot (X^I_{rs} \arccos(X^I_{rs}))\} (X^I_{rs} \nu_s) (X^I_{rs} \nu_s) + (1 - (X^I_{rs})^2)^{-1/2} \cdot (X^I_{rs} \arccos(X^I_{rs}))\} \delta_{rs}], 1 \leq r, s \leq d.
\end{equation}

Here $\{\nu_r : 1 \leq r \leq d\}$ is an orthonormal basis of $T_{\mu_I} S^d$. A confidence region for $\mu_I$, of asymptotic level $1 - \alpha$, is then given by
\begin{equation}
\{p \in S^d : n\phi(p(\mu_n))^t \hat{\Lambda}_p^{-1} \hat{\Lambda}_p \phi(p(\mu_n)) \leq \chi^2_d(1 - \alpha)\},
\end{equation}
where $\Lambda_p, \hat{\Sigma}_p$ are obtained by replacing $\mu_I$ by $p$ in the expressions for $\Lambda$ and $\hat{\Sigma}$ in (60). The 'hat' (') indicates that the expectations are computed under the empirical $Q_n$, rather than $Q$. As mentioned in Section 5, it would be computationally simpler to choose a particular $p = p_0$, say, and let $\phi = Exp_{\mu_I}^{-1}$. Then (42) yields a simpler confidence region:
\begin{equation}
\{p \in S^d : n[\phi(\mu_n) - \phi(p)]^t \hat{\Lambda}_p^{-1} \hat{\Lambda}_p [\phi(\mu_n)] - \phi(\mu_p)] \leq \chi^2_d(1 - \alpha)\}.
\end{equation}

Example 6.6. (Inference on the planar shape space $\Sigma^2_\hat{\gamma}$).

To apply Theorem 4.10, we use (54) where $\phi = Exp_{\gamma(p)}^{-1}$ and $p$ is a suitable point in $\mathbb{C}S^{k-1}$. To derive a computable expression for $\Lambda$, write the geodesic $\gamma$ in the parametric surface $c(s, t)$ as $\gamma = \pi_\gamma \hat{\gamma}$, where $\hat{\gamma}$ is a geodesic in $\mathbb{C}S^{k-1}$ starting at $\hat{\gamma}$. Then, with $\tilde{T}(s, 1) = (d\gamma(s) \pi^{-1}) T(s, 1),
\begin{equation}
(d/ds) \rho^2_\gamma(\gamma(s), m) = 2 < T(s, 1), \hat{\gamma}(s) > = 2 < \tilde{T}(s, 1), \hat{\gamma}(s) >,
\end{equation}
\begin{equation}
(d^2/ds^2) \rho^2_\gamma(\gamma(s), m) = 2 < D_\gamma \tilde{T}(s, 1), \hat{\gamma}(s) >.
\end{equation}
The final inner products are in $T\mathbb{C}S^{k-1}$, namely, $\langle \hat{v}, \hat{w} \rangle = \text{Re}(\hat{v}^* \hat{w})$. Note that $\tilde{T}(s, 1) = -\text{Exp}_{\hat{\gamma}}(s \bar{q}, q \in \pi^{-1}m$, may be expressed by (10) and (11) as
\begin{equation}
\tilde{T}(s, 1) = -\rho(s)/\sin \rho(s))] e^i\theta(s) q - (\cos \rho(s)) \hat{\gamma}(s),
\end{equation}
where $\rho(s) = \rho_{\gamma}(\gamma(s), m)$ and $e^i\theta(s) = (1/ \cos \rho(s)) \hat{\gamma}(s) q^*$. The covariant derivative $D_\gamma \tilde{T}(s, 1)$ is the projection of $(d/ds)\tilde{T}(s, 1)$ onto $H_{\gamma(s)}$. Since $\langle \hat{\mu}, \hat{\gamma}(0) \rangle = 0$, (63) then yields
\begin{equation}
[(d^2/ds^2) \rho^2_\gamma(\gamma(s), m)]_{s=0} = 2 \langle [d/ds] \tilde{T}(s, 1)]_{s=0}, \hat{\gamma}(0) \rangle.
\end{equation}
Differentiating (64) one obtains
\begin{equation}
[(d/ds) \tilde{T}(s, 1)]_{s=0} = [(d/ds)(\rho(s) \cos \rho(s))/\sin \rho(s)]_{s=0} \hat{\mu}
+ [(\rho(s) \cos \rho(s))/\sin \rho(s)]_{s=0} \hat{\gamma}(0) - [(d/ds)(\rho(s)/\cos \rho(s)) (\sin \rho(s))]_{s=0} (\hat{\mu} q^*) q
- [\rho(s)/\cos \rho(s)] (\sin \rho(s))]_{s=0} (\hat{\gamma}(0) q^*) q.
\end{equation}
From (63), \(2 \rho(s) \rho(s) = 2 \langle \tilde{T}(s, 1), \gamma'(s) \rangle\), which along with (64) leads to

\[
(67) \quad [(d/ds) \rho(s)]_{s=0} = - (1/sin(r))(\mu q^*/cos(r))(0), \quad (r = \rho_\theta(m, \mu_1)).
\]

One then gets (See Bhattacharya and Bhattacharya (2008), pp. 93, 94)

\[
(68) \quad [(d/ds) \tilde{T}(s, 1)]_{s=0}, \gamma'(0) = \{(r \cos r)/(\sin r)\}[(\gamma(0))\|^2
\]

\[- \{(1/sin^2(r) - (r \cos r)/\sin^3(r))(Re(x))^2 + r/((\sin r)(\cos r))(Im(x))^2, \]

\[
(x = e^{i0} q \gamma(0))^*, e^{i0} = \mu q^*/cos(r).
\]

One can check that the right side of (68) depends only on \(\pi(\tilde{\mu})\) and not any particular choice of \(\tilde{\mu}\) in \(\pi^{-1}\{\mu_1\}\).

Now let \(\{\nu_1, \cdots, \nu_{k-2}, i\nu_1, \cdots, i\nu_{k-2}\}\) be an orthonormal basis of \(T_{\sigma(p)} \Sigma_2^k\) where we identify \(\Sigma_2^k\) with \(\mathbb{C}P^{k-2}\), and choose the unit vectors \(\nu_r = (\nu_r^1, \cdots, \nu_r^{k-1})\), \(r = 1, \cdots, k-2\), to have zero imaginary parts and satisfy the conditions \(p^*\nu_r = 0, v_r^0 \nu_s = 0\) for \(r \neq s\).

Suppose now that \(\sigma(p) = \mu_1\), i.e., \(\gamma(0) = \mu_1\). If \(\gamma'(0) = v\), then \(\gamma(s) = \text{Exp}_{\mu_1}(sv)\), so that \(\rho_\sigma^2(\gamma(s), m) = h(sv, y)\) with \(y = \text{Exp}_{\mu_1}^{-1} m\). Then, expressing \(v\) in terms of the orthonormal basis,

\[
(69) \quad [(d^2/ds^2) \rho^2(\gamma(s), m)]_{s=0} = [(d^2/ds^2)h(sv, y)]_{s=0} = \Sigma v_i v_j D_i D_j h(0, y).
\]

Integrating with respect to \(Q\) now yields

\[
(70) \quad \sum v_i v_j A_{ij} = E[(d^2/ds^2) \rho^2(\gamma(s), \lambda)]_{s=0}. \quad (X\text{ with distribution } Q).
\]

This identifies the matrix \(A\) from the calculations (65) and (68). To be specific, consider independent observations \(X_1, \cdots, X_n\) from \(Q\), and let \(Y_j = \text{Exp}_{\mu_1}^{-1} X_j (j = 1, \cdots, n)\). In normal coordinates with respect to the above basis of \(T_{\mu_1, \Sigma_2^k}\), one has the following coordinates of \(Y_j\):

\[
(71) \quad (\text{Re}(Y_j^1), \cdots, \text{Re}(Y_j^{k-2}), \text{Im}(Y_j^1), \cdots, \text{Im}(Y_j^{k-2})) \in \mathbb{R}^{2k-4}.
\]

Writing

\[
\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}
\]

in blocks of \((k-2) \times (k-2)\) matrices, one arrives at the following expressions of the elements of these matrices, using (68)-(71). Denote \(\rho_\sigma^2(\mu_1, X_1) = h(0, \gamma_1)\) by \(\rho\). Then

\[
(72) \quad (\Lambda_{11})_{rs} = 2E[\rho(\cot \rho) \delta_{rs} - (1/\rho^2)(1 - \rho \cot \rho)(\text{Re} Y_1^r)(\text{Re} Y_1^s)
\]

\[
+ \rho^{-1}(\tan \rho)(\text{Im} Y_1^r)(\text{Im} Y_1^s)];
\]

\[
(\Lambda_{22})_{rs} = 2E[\rho(\cot \rho) \delta_{rs} - (1/\rho^2)(1 - \rho \cot \rho)(\text{Im} Y_1^r)(\text{Im} Y_1^s)
\]

\[
+ \rho^{-1}(\tan \rho)(\text{Re} Y_1^r)(\text{Re} Y_1^s)];
\]

\[
(\Lambda_{12})_{rs} = 2E[\rho(\cot \rho) \delta_{rs} - (1/\rho^2)(1 - \rho \cot \rho)(\text{Re} Y_1^r)(\text{Im} Y_1^s)
\]

\[
+ \rho^{-1}(\tan \rho)(\text{Im} Y_1^r)(\text{Re} Y_1^s)];
\]

\[
(\Lambda_{21})_{rs} = (\Lambda_{12})_{sr} (r, s = 1, \cdots, k - 2).
\]

One now arrives at the CLT for the intrinsic sample mean \(\mu_n\) by Theorem 4.10 and Corollary 4.4. A two-sample test for \(H_0 : Q_1 = Q_2\) is then provided by (42).

We next turn to extrinsic analysis on \(\Sigma_2^k\), using the embedding (12). Let \(\mu_J\) be the mean of \(Q \circ J^{-1}\) on \(S(k-1, \mathbb{C})\), where \(J\) is the veronese-Whitney embedding (12).
Assuming that the largest eigenvalue of \( \mu_J \) is simple (see proposition 3.1), one may now obtain the asymptotic distribution of the sample extrinsic mean \( \mu_{n,E} \), namely, that of 
\[
J(\mu_{n,E}) = \tilde{\nu}_n^* \nu_n,
\]
where \( \nu_n \) is a unit eigenvector of \( \tilde{Y} = \sum \hat{Y}_j/n \) corresponding to its largest eigenvalue. Here \( \hat{Y}_j = J(Y_j) \), for i.i.d observations \( Y_1, \ldots, Y_n \) on \( \Sigma_k^2 \). For this purpose, a convenient orthonormal basis (frame) of \( T_pS(k-1, \mathbb{C}) \approx S(k-1, \mathbb{C}) \) is the following:
\[
\begin{align*}
\nu_{a,b} & = 2^{-1/2}(e_a e_b^t + e_b e_a^t) \quad \text{for} \ a < b, \ \nu_{a,a} = e_a e_a^t; \\
w_{a,b} & = i2^{-1/2}(e_a e_b^t - e_b e_a^t) \quad \text{for} \ b < a \ (a, b = 1, \ldots, k-1),
\end{align*}
\]
where \( e_a \) is the column vector with all entries zero other than the \( a \)-th, and the \( a \)-th entry is 1. Let \( U_1, \ldots, U_{k-1} \) be orthonormal unit eigenvectors corresponding to the eigenvalues \( \lambda_1 \leq \cdots \leq \lambda_{k-2} < \lambda_{k-1} \). Then choosing \( T = (U_1, \ldots, U_{k-1}) \in SU(k-1) \)
\[
T \mu_J T^* = D = \text{diag}(\lambda_1, \ldots, \lambda_{k-1})
\]
such that the columns of \( TV_a T^* \) and \( Tw_{a,b} T^* \) together constitute an orthonormal basis of \( S(k-1, \mathbb{C}) \). It is not difficult to check that the differential of the projection operator \( P \) satisfies
\[
(d_P \mu)Tv_{a,b} T^* = \begin{cases} 
0 & \text{if} \ 1 \leq a < b < k-1, \ \text{or} \ a = b = k-1, \\
(\lambda_{k-1} - \lambda_a)^{-1}Tv_{a,k-1}T^* & \text{if} \ 1 \leq a < k-1, b = k-1;
\end{cases}
\]
\[
(d_P \mu)Tw_{a,b} T^* = \begin{cases} 
0 & \text{if} \ 1 \leq a \leq b < k-1, \\
(\lambda_{k-1} - \lambda_a)^{-1}Tw_{a,k-1}T^* & \text{if} \ 1 \leq a < k-1.
\end{cases}
\]
To check these, take the projection of a linear curve \( c(s) \) in \( S(k-1, \mathbb{C}) \) such that \( c(0) \) is one of the basis elements \( v_{a,b} \), or \( w_{a,b} \), and differentiate the projected curve with respect to \( s \). It follows that \( \{Tv_{a,k-1} T^*, Tw_{a,k-1} T^* : a = 1, \ldots, k-2 \} \) form an orthonormal basis of \( TP(\mu_J)J(\Sigma_k^2) \). Expressing \( \tilde{Y}_j - \mu_J \) in the orthonormal basis of \( S(k-1, \mathbb{C}) \), and \( d_P \mu (\tilde{Y}_j - \mu_J) \) with respect to the above basis of \( TP(\mu_J)J(\Sigma_k^2) \), one may now apply Proposition 4.3.

For a two-sample test for \( H_0 : Q_1 = Q_2 \), one may use (44), as explained in Section 5.

7. Nonparametric Bayes Estimation of Densities on a Manifold and the Problem of Classification.

7.1. Density estimation. Consider the problem of estimating the density \( q \) of a distribution \( Q \) on a Riemannian manifold \( (M, g) \) with respect to the volume measure \( \lambda \) on \( M \). According to Ferguson (1973), given a finite non-zero base measure \( \alpha \) on a measurable space \( (\mathcal{X}, \Sigma) \), a random probability \( P \) on the class \( \mathcal{P} \) of all probability measures on \( \mathcal{X} \) has the Dirichlet distribution \( D_\alpha \) if for every measurable partition \( \{B_1, \ldots, B_k\} \) of \( \mathcal{X} \), the \( D_\alpha \) - distribution of \( (P(B_1), \ldots, P(B_k)) = (\theta_1, \ldots, \theta_k) \), say, is Dirichlet with parameters \( (\alpha(B_1), \ldots, \alpha(B_k)) \). That is, \( (P(B_1), \ldots, P(B_k)) \) has the joint density
\[
f(\theta_1, \ldots, \theta_k) = \text{const} \left( \theta_1^{\alpha(B_1)-1} \cdots \theta_k^{\alpha(B_k)-1} \right) (1 - \theta_1 - \cdots - \theta_k)^{\alpha(B_k)-1}
\]
on \( \{(\theta_1, \ldots, \theta_k) : \theta_i > 0 \forall i, \ \theta_1 + \cdots + \theta_k < 1 \} \). If \( \alpha(B) = 0 \) for some \( B_j \), then \( P(B_j) = 0 \) with probability 1. In the case \( k = 2 \), the \( D_\alpha \)-distribution of \( (P(B_1), P(B_2)) \) is also called the beta distribution, denoted beta(\( \alpha(B_1), \alpha(B_2) \)). Sethuraman (1994) gave a very convenient “stick breaking” representation of the random \( P \). To define it, let \( u_j (j = 1, \ldots) \) be an i.i.d. sequence of beta(1, \( \alpha(\mathcal{X}) \)) random variables, independent of a sequence
\( Y_j (j = 1, \ldots) \) having the distribution \( G = \frac{p_j}{\alpha(X)} \) on \( X \). Sethuraman’s representation of the random probability with the Dirichlet prior distribution \( D_\alpha \) is

\[
P \equiv \sum w_j \delta_{Y_j},
\]

where \( w_1 = u_1, w_j = u_j (1 - u_1) \ldots (1 - u_{j-1}) (j = 2, \ldots), \) and \( \delta_{Y_j} \) denotes the Dirac measure at \( Y_j \). As this construction shows, the Dirichlet distribution assigns probability one to the set of all discrete distributions on \( X \), and one cannot retrieve a density estimate from it directly. The Dirichlet priors constitute a conjugate family, i.e., the posterior distribution of a random \( P \) with distribution \( D_\beta \), given observations \( X_1, \ldots, X_n \) from \( P \) is \( D_\alpha + \sum_{i \leq n} \delta_{X_i} \). A general method for Bayesian density estimation on a manifold \((M, g)\) may be outlined as follows. Suppose that \( q \) is continuous and positive on \( M \). First find a parametric family of densities \( m \to K(m; \mu, \tau) \) on \( M \) where \( \mu \in M \) and \( \tau > 0 \) are “location” and “scale” parameters, such that \( K \) is continuous in its arguments, \( K(\cdot; \mu, \tau) d\lambda(\cdot) \) converges to \( \delta_\mu \) as \( \tau \to 0 \), and the set of all “mixtures” of \( K(\cdot; \mu, \tau) \) by distributions on \( M \times (0, \infty) \) is dense in the set \( C_\lambda(M) \) of all continuous densities on \( M \) in the supremum distance, or in \( L^1(d\lambda) \). The density \( q \) may then be estimated by a suitable mixture. To estimate the mixture, use a prior \( D_\beta \) with full support on the set of all probabilities on the space \( M \times (0, \infty) \) of “parameters” \((\mu, \tau)\). A draw from the prior may be expressed in the form (75), where \( u_j \) are i.i.d. \( beta(1, b) \) with \( b = \beta(M \times (0, \infty)) \), independent of \( Y_j = (m_j, t_j) \), say, which are i.i.d. \( \frac{b}{b+n} \) on \( M \times (0, \infty) \). The corresponding random density is then obtained by integrating the kernel \( K \) with respect to this random mixture distribution,

\[
\sum w_j K(m; m_j, t_j).
\]

7.2. **Classification.** Classification of a random observation to one of several groups is one of the most important problems in statistics. This is the objective in medical diagnostics, classification of subspecies and, more generally, this is the target of most image analysis. Suppose there are \( r \) groups or populations with a priori given relative sizes or proportions \( \pi_i (i = 1, \ldots, r), \sum \pi_i = 1 \), and densities \( q_i(x) \) (with respect to some sigma-finite measure). Under \( 0-1 \) loss function, the average risk of misclassification (i.e., the Bayes risk)
is minimized by the rule: Given a random observation $X$, classify it to belong to group
$j$ if $\pi_j q_j(X) = \max\{\pi_i q_i(X) : i = 1, \ldots, r\}$. Generally, one uses sample estimates of
$\pi_i$-s and $q_i$-s, based on random samples from the $r$ groups (training data). Nonparametric
Bayes estimates of $q_i$-s on shapes spaces perform very well in classification of shapes,
and occasionally identify outliers and misclassified observations (See, Bhattacharya and
Bhattacharya (2012) and Bhattacharya and Dunson (2010)).

In a simulation study using 20 random draws from a complex Watson distribution, Bhat-
tacharya and Dunson (2010) found that the nonparametric Bayes estimate far outperformed
the kernel density estimate KDE over a multitude of criteria. It also performed much bet-
ter than the MLE of the correctly specified model! Here are the $L^1$ distances and the
Kullback-Leibler divergences from $f_0$.

|              | NP Bayes | KDE  | MLE  |
|--------------|----------|------|------|
| $L^1$        | 0.44     | 1.03 | 0.75 |
| K-L          | 0.13     | 0.41 | 0.25 |

8. THE LAPLACE-BELTRAMI OPERATOR IN MACHINE VISION

Mark Kac asked in a paper in 1966 in the American Mathematical Monthly: “Can one
hear the shape of a drum?” In other words, by listening to the frequencies of vibrations of a
clamped drum, given by the eigenvalues of the Laplacian with Dirichlet (or zero) boundary
condition, is it possible to reconstruct or identify the geometric shape of the drum? The
origin of this question may be traced back to Hermann Weyl’s famous formula (Weyl,
1911): For any bounded domain $\Omega$ in $\mathbb{R}^d$ with a smooth boundary, the number $N(\lambda)$
of eigenvalues of the (negative) Laplacian $-\Delta$ which are less than $\lambda$ has the asymptotic relation

$$N(\lambda) \sim \omega_d (2\pi)^{-d/2} \lambda^{d/2} \text{vol}(\Omega) \text{ as } \lambda \to \infty \quad (\omega_d = \text{vol of unit ball in } \mathbb{R}^d).$$

Here the relation $\sim$ indicates that the ratio of its two sides converges to 1 (as $\lambda \to \infty$). A
similar formula holds for any $d$-dimensional compact Riemannian manifold $(M, g)$ with
or without boundary where $\Delta$ is the so-called Laplace Beltrami operator (Chavel, 1984;
Rosenberg, 1997), which may be expressed in a local chart given by $u$ (on $B(0, r) \to
U \subset M$) as

$$\Delta f = (\det g)^{-1/2} \sum_{1 \leq i, j \leq d} \partial_i g^{ij} (\det g)^{1/2} \partial_j f.$$  

Weyl type spectral asymptotics for $-\Delta$ are given by

$$N(\lambda) \sim c(d) \lambda^{d/2} \text{vol}(\Omega) \text{ as } \lambda \to \infty, \text{ where } c(d) \text{ depends only on the dimension } d.$$  

There are many refinements of the estimates (77), (32) with an error term of the order $\lambda^{(d-1)/2}$. Although there are many spectral invariants of the manifold, it turns out unfortunately, that the answer to Kac’s question is “no” (Milnor, 1964). For two-dimensional
surfaces, the answer is mostly “yes” outside a relatively small set of manifolds (Zelditch,
2000). But in dimension 3 or higher the set of non-isometric manifolds with the same spec-
trum is not negligible. A natural question that arises is: if one uses eigenfunctions as well
as eigenvalues of $-\triangle$ can one reconstruct or identify the manifold? One may find many interesting and important articles in computer science/machine vision journals where such reconstructions are displayed. But the mathematical question posed above is rigorously answered in the affirmative only by Jones et al. (2008), under only mild conditions on the manifold, such as uniform ellipticity of the Laplacian. The last mentioned authors actually construct coordinate patches covering $M$, and therefore the structure of the manifold, only using eigenvalues and eigenfunctions of $-\triangle$.

There are many numerical methods for the computation of eigenvalues and eigenfunctions carried out by computer scientists and applied for object identification and scene recognition (See, e.g. Reuter (2006)). To economize the use of these features sometimes topological properties of $M$ are also used. For example, see Dey et al. (2008), Dey and Li (2009) who use the first homology group to identify handles and holes in a closed bounded domain in 3D. For a more elaborate technique using algebraic topology, known as persistent homology, we refer to Carlsson (2009).

9. Examples and Applications

In this section we apply the theory to a number of data sets available in the literature.

**Example 9.1.** (Paleomagnetism). The first statistical confirmation of the shifting of the earth’s magnetic poles over geological times, theorized by paleontologists based on observed fossilised magnetic rock samples, came in a seminal paper by R.A. Fisher (1953). Fisher analyzed two sets of data - one recent (1947-48) and another old (Quaternary period), using the so-called von Mises-Fisher model

$$f(x; \mu, \tau) = c(\tau) \exp\{\tau x^t \mu\} (x \in S^2),$$

Here $\mu (\in S^2)$, is the mean direction, extrinsic as well as intrinsic ($\mu = \mu_I = \mu_E$), and $\tau > 0$ is the concentration parameter. The maximum likelihood estimate of $\mu$ is $\hat{\mu} = \bar{X}/|\bar{X}|$, which is also our sample extrinsic mean. The value of the MLE for the first data set of $n = 9$ observations turned out to be $\hat{\mu} = \hat{\mu}_E = (0.2984, 1.346, .9449)$, where $(0,0,1)$ is the geographic north pole. Fisher’s 95% confidence region for $\mu$ is $\{\mu \in S^2 : \rho_g(\hat{\mu}, \mu) \leq 0.1536\}$. The sample intrinsic mean is $\hat{\mu}_I = (0.2990, 0.349, 0.9447)$, which is very close to $\hat{\mu}_E$. The nonparametric confidence region based on $\hat{\mu}_I$, as given by (61), and that based on the extrinsic procedure (44), are nearly the same, and both are about 10% smaller in area than Fisher’s region. (See Bhattacharya and Bhattacharya (2012), Chapter 2).

The second data set based on $n = 29$ observations from the Quaternary period that Fisher analyzed, using the same parametric model as above, had the MLE $\hat{\mu} = \bar{X}/|\bar{X}| = (0.0172, -.2978, -.9545)$, almost antipodal of that for the first data set, and with a confidence region of geodesic radius .1475 around the MLE. Note that the two confidence regions are not only disjoint, they also lie far away from each other. This provided the first statistical confirmation of the hypothesis of shifts in the earth’s magnetic poles, a result hailed by paleontologists (See Irving (1964)). Because of difficulty in accessing the second data set, the nonparametric procedures could not be applied to it. But the analysis of another data set dating from the Jurassic period, with $n = 33$, once again yielded nonparametric intrinsic and extrinsic confidence regions very close to each other, and each about 10% smaller than the region obtained by Fisher’s parametric method (See Bhattacharya and Bhattacharya (2012), Chapter 5, for details).
Example 9.2. (Brain scan of schizophrenic and normal patients). We consider an example from Bookstein (1991) in which 13 landmarks were recorded on a midsagittal two-dimensional slice from magnetic brain scans of each of 14 schizophrenic patients and 14 normal patients. The object is to detect the deformation, if any, in the shape of the k-ad due to the disease, and to use it for diagnostic purposes. The shape space is $\Sigma_{13}^2$. The intrinsic two-sample test (46) has an observed value 95.4587 of the asymptotic chi-square statistic with 22 degrees of freedom, and a $p$-value $3.97 \times 10^{-11}$. The extrinsic test based on (46) has an observed value 95.5476 of the chi-square statistic and a $p$-value $3.8 \times 10^{-11}$. It is remarkable, and reassuring, that completely different methodologies of intrinsic and extrinsic inference essentially led to the same values of the corresponding asymptotic chi-square statistics (a phenomenon observed in other examples as well). For details of these calculations and others we refer to Bhattacharya and Bhattacharya (2012). This may also be contrasted with the results of parametric inference in the literature for the same data, as may be found in Dryden and Mardia (1998), pp. 146, 162-165. Using a isotropic Normal model for the original landmarks data, and after removal of "nuisance" parameters for translation, size and rotation, an $F$-test known as Goodall’s $F$-test (See Bookstein (1991)) gives a $p$-value .01. A Monte Carlo test based permutation test obtained by 999 random assignments of the data into two groups and computing Goodall’s $F$-statistic, gave a $p$-value .04. A Hotelling’s $T^2$ test in the tangent space of the pooled sample mean had a $p$-value .834. A likelihood ratio test based on the isotropic offset Normal distribution on the shape space has the value 43.124 of the chi-square statistic with 22 degrees of freedom, and a $p$-value .005.

Example 9.3. (Shapes of Gorilla Skulls)

We consider another example in which two planar shape distributions via their extrinsic (and intrinsic) means are distinguished. A Bayesian nonparametric classifier is also built and applied.

In this data set, there are 29 male and 30 female gorillas and the eight landmarks are chosen on the midline plane of the 2D image of the skull. The data can be found in Dryden and Mardia (1998). It is of interest to study the shapes of the skulls and use that to detect differences in shapes between the sexes. This finds applications in morphometrics and other biological sciences.

To distinguish between the distribution of shapes of skulls of the two sexes, one may compare the sample extrinsic mean shapes or dispersions in shape as well as the intrinsic counterparts.

The value of the two sample test statistic defined in (46), for comparing the intrinsic mean shapes, and the asymptotic $p$-value for the chi-squared test are

$$T_{n1} = 391.63, \text{ p-value } = P(\chi^2_{12} > 391.63) < 10^{-16}.$$ 

Hence we reject the null hypothesis that the two sexes have the same intrinsic mean shape. The test statistics, defined in equations (44) for comparing the extrinsic mean shapes, and the corresponding asymptotic $p$-values are

$$T_1 = 392.6, \text{ p-value } = P(\chi^2_{12} > 392.6) < 10^{-16}.$$ 

Hence we reject the null hypothesis that the two sexes have the same extrinsic mean shape. We can also compare the mean shapes by pivotal bootstrap method using the test statistic $T_2^*$ which is a bootstrap version of $T_2$. The $p$-value for the bootstrap test using $10^5$ simulations turns out to be 0. In contrast, a parametric test carried out in Dryden and Mardia (1998), pp. 168-172, has a $p$-value .0001.
A Bayesian nonparametric classifier is next applied (see Bhattacharya and Dunson (2010)) to predict gender. The shape densities for the two groups via non-parametric Bayesian methods are estimated which are used to derive the conditional distribution of gender given shape. 25 individuals of each gender are picked as a training sample, with the remaining 9 used as test data. Table 2 presents the estimated posterior probabilities of being female for each of the gorillas in the test sample along with a 95% credible interval. For most of the gorillas, there is a high posterior probability of assigning the correct gender. There is misclassification only in the 3rd female and 3rd male. For the 3rd female, the credible interval includes 0.5, suggesting that there is insufficient information to be confident in the classification. However, for the 3rd male, the credible interval suggests a high degree of confidence that this individual is female. Perhaps this individual is an outlier and there is something unusual about the shape of his skull, with such characteristics not represented in the training data, or, alternatively, he was labeled incorrectly.

**Example 9.4** (Corpus Callosum shapes of normal and ADHD children)

We consider the third planar shape data set, which involve measurements of a group typically developing children and a group of children suffering the ADHD (Attention deficit hyperactivity disorder). ADHD is one of the most common psychiatric disorders for children that can continue through adolescence and adulthood. Symptoms include difficulty staying focused and paying attention, difficulty controlling behavior, and hyperactivity (over-activity). ADHD in general has three subtypes: (1) ADHD hyperactive-impulsive (2) ADHD-inattentive; (3) Combined hyperactive-impulsive and inattentive (ADHD-combined). Ramsay (2007). ADHD-200 Dataset (http://fcon_1000.projects.nitrc.org/indi/adhd200/) is a data set that record both anatomical and resting-state functional MRI data of 776 labeled subjects across 8 independent imaging sites, 491 of which were obtained from typically developing individuals and 285 in children and adolescents with ADHD (ages: 7-21 years old). The Corpus Callosum shape data are extracted using the CCseg package, which contains 50 landmarks on the contour of the Corpus Callosum of each subject (see Huang et al. (2015)). After quality control, 647 CC shape data out of 776
subjects were obtained, which included 404 \((n_1)\) typically developing children, 150 \((n_2)\) diagnosed with ADHD-Combined, 8 \((n_3)\) diagnosed with ADHD-Hyperactive-Impulsive, and 85 \((n_4)\) diagnosed with ADHD-Inattentive. Therefore, the data lie in the space \(\Sigma_2^{50}\), which has a high dimension of \(2 \times 50 - 4 = 96\).

We carry out extrinsic two sample tests between the group of typically developing children and the group of children diagnosed with ADHD-Combined, and also between the group of typically developing children and ADHD-Inattentive children. We construct test statistics that base on the asymptotic distribution of the extrinsic mean for the planar shapes.

The \(p\)-value for the two-sample test between the group of typically developing children and the group of children diagnosed with ADHD-Combined is \(5.1988 \times 10^{-11}\), which is based on the asymptotic chi-squared distribution given in (44). The \(p\)-value for the test between the group of typically developing children and the group ADHD-Inattentive children is smaller than \(10^{-50}\).

**Example 9.5** (Positive definite matrices with application to diffusion tensor imaging.) Another important class of manifolds is \(\text{sym}^+(p)\), the space of \(p \times p\) positive definite matrices. In particular, when \(p = 3\), \(\text{sym}^+(3)\), the space of \(3 \times 3\) positive definite matrices, has important applications in diffusion tensor imaging (DTI). DTI is now an important tool for neuroimaging in clinical trials. It provides for the measurement of the diffusion matrix \((3 \times 3\) positive definite matrice) of molecules of water in tiny voxels in the white matter of the brain. When there are no barriers, the diffusion matrix is isotropic, and in the presence of structural barriers in the brain white matter due to axon (nerve fiber) bundles and their myelin sheaths (electrically insulating layers) the diffusion is anisotropic, and DTI can be used to measure the anisotropic diffusion tensor. When a trauma occurs, due to an injury or a disease, this highly organized structure is disrupted and anisotropy decreases. Large scale DTI based studies have been used to investigate autism, schizophrenia, Parkinson’s disease and Alzheimer’s disease. The geometry of \(\text{sym}^+(p)\) for general \(p\) is now described in the following.

Let \(A \in \text{sym}^+(p)\) which follows a distribution \(Q\). We first introduce the Euclidean metric of \(A\), which is given by \(\|A\|^2 = \text{Trace}(A)^2\). Since \(\text{sym}^+(p)\) is an open convex subset of \(\text{sym}(p)\), the space of all \(p \times p\) symmetric matrices, the mean of \(Q\) with respect to the Euclidean distance is given by the Euclidean mean

\[
\mu_E = \int A Q(dA).
\]  

Another important metric for \(\text{sym}^+(p)\) is the log-Euclidean metric (Arsigny et al., 2006). Let \(J \equiv \log : \text{sym}^+(p) \to \text{sym}(p)\) be the inverse of the exponential map \(B \to e^B\), \(\text{sym}(p) \to \text{sym}^+(p)\), which is the matrix exponential of \(B\). Then \(J\) is a diffeomorphism. The log Euclidean distance is given by

\[
\rho_{LE}(A_1, A_2) = \| \log(A_1) - \log(A_2) \|.
\]  

Note that \(J\) is an embedding on \(\text{sym}^+(p)\) onto \(\text{sym}(p)\) and, in fact, it is an equivariant embedding under the group action of \(\text{GL}(p, \mathbb{R})\), the general linear group of \(p \times p\) non-singular matrices. The extrinsic mean of \(Q\) under \(J\) is given by

\[
\mu_{E,J} = \exp\left( \int (\log(A)) Q(dA) \right).
\]  

We apply Theorem 4.2 to sample Fréchet means under both the Euclidean and log-Euclidean distances. In particular, we consider a diffusion tensor imaging (DTI) data set consisting of 46 subjects with 28 HIV+ subjects and 18 healthy controls. Diffusion tensors
were extracted along the fiber tract of the splenium of the corpus callosum. The DTI data for all the subjects are registered in the same atlas space based on arc lengths, with 75 features obtained along the fiber tract of each subject. This data set has been studied in a regression setting in Yuan et al. (2012). On the other hand, we carry out two sample tests between the control group and the HIV+ group for each of the 75 sample points along the fiber tract. Therefore, 75 tests are performed in total. Two types of tests are carried out based on the Euclidean distance and the log-Euclidean distance.

The simple Bonferroni procedure for testing $H_0$ yields a $p$-value equal to 75 times the smallest $p$-value which is of order $10^{-7}$. To identify sites with significant differences, the 75 $p$-values are ordered from the smallest to the largest with a false discovery rate of $\alpha = 0.05$. 58 sites are found to yield significant differences using the Euclidean distance, and 47 using the log-Euclidean distance (see Benjamini and Hochberg (1995)).

**Example 9.6.** (Glaucoma detection- a match pair problem in 3D). Our final example is on the 3D reflection similarity shape space $R\Sigma^k$. To detect shape changes due to glaucoma, data were collected on twelve mature rhesus monkeys. One of the eyes of each monkey was treated with a chemical agent to temporarily increase the intraocular pressure (IOP). The increase in IOP is known to be a cause of glaucoma. The other eye was left untreated. Measurements were made of five landmarks in each eye, suggested by medical professionals. The data may be found in Bhattacharya and Patrangenaru (2005). The match pair test based on (47) yielded an observed value 36.29 of the asymptotic chisquare statistic with degrees of freedom 8. The corresponding $p$-value is $1.55 \times 10^{-5}$ (See Bhattacharya and Bhattacharya (2012), Chapter 9). This provides a strong justification for using shape change of the inner eye as a diagnostic tool to detect the onset of glaucoma. An earlier computation using a different nonparametric procedure in Bhattacharya and Patrangenaru (2005) provided a $p$-value .058. Also see Bandulasiri et al. (2009) where a 95% confidence region is obtained for the difference between the extrinsic size-and-shape reflection shapes between the treated and untreated eyes.

**APPENDIX A. APPENDIX ON RIEMANNIAN MANIFOLDS**

Often the manifold $M$ in applications has a natural Riemannian metric tensor $g$. That is, it is given an inner product $\langle \cdot, \cdot \rangle_p$ on the tangent space $T_p M$ at $p$, which is smoothly defined. In local coordinates in $U_p$ given by $\psi_p(\cdot) = x = (x_1, \ldots, x_d) \in B_p$, the functions $(g_{ij})(x) = (E_i, E_j)_p$, with $E_i = d\psi_p^1(\partial/\partial x_i)$ ($i, j = 1, \ldots, d$), are smooth in $B_p$. This allows one to measure the length of a smooth arc $\gamma$ joining any two points $q, q'$ in $U_p$, namely, $\int_{[a,b]} |dx(t)/dt|dt, \gamma(a) = q, \gamma(b) = q', x(t) = \psi_p \circ \gamma(t)$. Here $|dx(t)/dt|^2 = \langle dx(t)/dt, dx(t)/dt \rangle_p$, with $dx(t)/dt$ expressed in the local frame $E_i(i = 1, \ldots, d)$. One may also write $dx(t)/dt$ as $d\gamma(t)/dt$. Using the compatibility condition (ii) above one now defines the length of a smooth arc joining any two points in $M$. The geodesic distance $\rho_g(p, q)$ between $p$ and $q$ is the minimum of lengths of all smooth arcs joining $p$ and $q$.

A standard parametrization of a curve is its arc length $s$: $s = \int_{[a,b]} |d\gamma(u)/du|du$. In this parametrization of curves, one has $|d\gamma(t)/dt| = 1$. We will adopt this so called unit speed parametrization unless otherwise specified. The property of local minimization of arc lengths yields a first order condition on the velocity $d\gamma(t)/dt$ of the minimizing curve $\gamma$ at $t$: the acceleration along $\gamma$ is zero at every parameter join $t$. If $M$ is a submanifold (hyper surface) of an Euclidean space $\mathbb{R}^N$, then the second derivative $d^2\gamma(t)/dt^2$ is well defined, but in general does not belong to the tangent space of $M$ at $\gamma(t)$. By ‘acceleration’ one means the orthogonal projection of the vector $d^2\gamma(t)/dt^2$ onto the tangent space of $M$ at $\gamma(t)$. This projection is called the covariant derivative of the velocity and denoted...
The “zero acceleration” of a geodesic \( \gamma \) means \((D/dt)d\gamma(t)/dt = 0\). On a general differentiable manifold, which is not given explicitly as a submanifold, there is no “outside”. The proper extension of the above notion of covariant derivative by Levi-Civita, using a notion known as affine connection, for all differentiable manifolds was a milestone in the development of differential geometry (See, e.g., Do Carmo (1992) Chapter 2).

In local coordinates the equation for a geodesic is a second order ordinary differential equation. By the standard existence theorem for ordinary differential equations, a geodesic \( \gamma \) is uniquely determined on a maximal interval \((a, b)\) \((-\infty \leq a < b < \infty\)), given an initial point \( \gamma(0) = p \) and an velocity \( (d\gamma(t)/dt)_{t=0} = v \). According to a result of Hopf and Rinow (Do Carmo (1992), Chapter 7), the geodesics can be extended indefinitely, (i.e., \(a = -\infty \) and \( b = \infty \)), i.e., it is geodesically complete, if and only if \((M, \rho_g)\) is a complete metric space; this in turn is equivalent to the topological condition (2). In particular, all compact Riemannian manifolds are geodesically complete. In most of the applications in this article \( M \) is compact.

On a complete Riemannian manifold, a geodesic \( \gamma(t) = \gamma(t; p, v) \), \( t \geq 0 \), in the direction \( v \), is completely determined by an initial point \( p = \gamma(0) \), and an initial velocity \( v = (d\gamma(t)/dt)t = 0 \). A cut point of \( p \) of the geodesic \( \gamma \) along \( v \) is \( r(v) \) (or \( r(v); p, v \)), where \( r(v) \) is the supremum of all \( t_0 \) such that \( \gamma \) is distance minimizing between \( p = \gamma(0) \) and \( \gamma(t_0) \). The set of all cut points (along all \( v \)) is called the cut locus of \( p \), denoted \( \text{Cut}(p) \). The geodesic distance \( q \to \rho_g(p, q) \) may not be smooth at the cut locus \( \text{Cut}(p) \), as Example A.1 below shows. Next, define the exponential function \( \text{Exp}_p : T_p(M) \to M : \text{Exp}_p(v) = \gamma(1; p, v) \) the point in \( M \) reached by the geodesic in time \( t = 1 \), starting at \( p \) with an initial velocity \( v \). It is known that \( \text{Exp}_p \) is a diffeomorphism on an open ball \( B(0 : r_0) \) of \( T_p(M) \), of radius \( r_0 = r_0(p) < \infty \), onto \( M \setminus \text{Cut}(p) \) (Do Carmo (1992), p. 271). Here \( r_0 = r_0(p) \) is the geodesic distance between \( p \) and \( \text{Cut}(p) \). The inverse map \( \text{Exp}_p^{-1} : M \setminus \text{Cut}(p) \to T_p(B(0 : r_0)) \) is called the inverse exponential, or the log map, \( \log_p \), at \( p \). The quantity \( \text{inj}(M) = \sup\{r_0(p); p \in M\} \) is the injectivity radius of \( M \). The log map also provides the so called normal coordinates for a neighborhood of \( p \).

**Example A.1 (Exponential and Log Maps on the Sphere \( S^d \)).** Consider the unit sphere \( S^d = \{ x \in \mathbb{R}^{d+1} : |x|^2 = 1 \} \). Because \( |\gamma(t)| = 1 \) \( \forall t \) for a curve on \( S^d \), the tangent space at \( p \) may be identified as the set of vectors in \( \mathbb{R}^{d+1} \) orthogonal to \( p \), \( T_p(S^d) = \{ v \in \mathbb{R}^{d+1} : pv' = 0 \} \). Here we write \( p, v, v' \), etc. as row vectors. The geodesics are the big circles, so that the point reached at time one by the geodesic from \( p \) moving with an initial velocity \( v \) is the point on the big circle lying on the plane spanned by \( p \) and \( v \) at an arc distance \( |v| \), i.e.,

\[
\text{Exp}_p(v) = \cos(|v|)p + \sin(|v|)v/|v|, \quad v \neq 0, \quad \text{Exp}_p(0) = p (pv' = 0).
\]

Also, the geodesic distance between \( p \) and \( q \) is the smaller of the lengths \(|v|\) of the two arcs joining \( p \) and \( q \) on the big circle,

\[
\rho_g(p, q) = \arccos p q' \in [0, \pi].
\]

Note that the cut locus of \( p \) is \( \text{Cut}(p) = \{-p\} \), and the distance between \( p \) and \( -p \) is \( \pi \), and \( \text{inj}(S^d) = \pi \). Hence the map \( \log_p(q) \) is defined on \( S^d \setminus \{-p\} \) and obtained by solving for \( v \) the equation \( \text{exp}_p(v) = q \). Now \( |v| = \rho_g(p, q) \). Plugging this in (A.1) and using (A.2), one has

\[
\sqrt{1 - (pq')^2}v = [q - (pq')p]\arccos pq', \quad (p \neq q),
\]
which yields
\[
\log_p(q) = |q - (pq')p|(\log \cos p'q)/\sqrt{1 - (pq')^2}
\]
\[(A.3) = [\log_p(p, q)/\sin \log_p(p, q)]|q - (pq')p|,
\]
for \(q \neq p, q \neq -p, \log_p(p) = 0.\) The map \(\log_p(q)\) is a diffeomorphism on \(S^d \setminus \{-p\}\) onto \(\{v \in T_pS^d : |v| < \pi\}.\) If one uses complex coordinates for \(p, q\) then \(pq'\) in the formula above are to be replaced by \(Re(pq^*), etc.\)

Most of the manifolds we consider in this article are of the form \(M = N/G.\) Here \(N\) is a complete Riemannian manifold with a metric tensor \(\rho_{g,N}\) and \(G\) is a compact Lie group of isometries acting freely on \(N, i.e.,\) except for the identity map, no \(g\) in \(G\) has a fixed point. This means that the orbit \(O_p\) of a point \(p\) under \(G\) is in one-one correspondence with \(G.\) As a subset of \(N, O_p\) is a submanifold of \(N\) of dimension that of \(G.\) Its tangent space \(T_pO_p\) as a subspace of \(T_pN\) is called the vertical subspace of \(T_pN,\) denoted \(V_p.\) The subspace \(H_p,\) of \(T_pN\) orthogonal to \(V_p\) is the horizontal subspace. \(M\) is then a Riemannian manifold with the metric tensor. The projection \(\pi : N \to M\) is a Riemannian submersion. The quotient \(N/G\) is then a Riemannian manifold.

The final important notion from geometry needed in this section is that of curvature. First, consider a smooth unit speed curve \(\gamma\) in \(\mathbb{R}^2:\) \(1 = |\dot{\gamma}(t)|^2 = (\dot{\gamma}(t), \dot{\gamma}(t)).\) Differentiation shows that \(\ddot{\gamma}(t) = d\gamma(t)/dt\) is orthogonal to \(\dot{\gamma}(t) : \ddot{\gamma}(t) = \kappa(t)N(t),\) where \(N(t)\) is a unit vector orthogonal to \(\dot{\gamma}(t)\) such that \((\dot{\gamma}(t), N(t))\) has the same orientation as \((\partial/\partial x_1, \partial/\partial x_2).\) Then \(\kappa(t)\) is the curvature of \(\gamma\) at the point \(\gamma(t).\) Next, at a point \(p\) on a regular surface \(S\) in \(\mathbb{R}^3,\) let \(N = \gamma(p)\) denote a unit normal to \(S\) at \(p.\) A plane \(\pi\) through \(N(p)\) intersects \(S\) in a smooth curve. Let \(\kappa(\cdot; p, \pi)\) be the curvature of this curve. As \(\pi\) varies by degrees of rotation, the curvature varies. Let \(\kappa_1\) be the maximum and \(\kappa_2\) the minimum of these curvatures, and let \(\kappa = \kappa_1\kappa_2.\) The Theorem Egregium of Gauss says that \(\kappa = \kappa(p)\) \((p \in S),\) the so-called Gaussian curvature, is intrinsic to the surface \(S,\) i.e., it is the same for all surfaces isometric to \(S\) (See, e.g., Boothby (1986), pp. 377-381). We now consider, somewhat informally, the case of a Riemannian manifold \(M.\) For \(p \in M\) and \(u, v \in T_p(M),\) consider the two dimensional subspace \(\pi\) spanned by \(u, v.\) Consider the two-dimensional submanifold swept out by geodesics in \(M\) with initial velocities lying in this subspace. The Gaussian curvature of this submanifold, thought of locally as a surface, is called the sectional curvature of \(M\) at \(p\) for the section \(\pi.\)

ACKNOWLEDGEMENT

The authors acknowledge support for this article from NSF grants DMS 1406872, CAREER 1654579 and IIS 1663870.

REFERENCES

Afsari, B. (2011). Riemannian \(L^p\) center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673.

Aggarwal, G., Chowdhury, A., and Chellappa, R. (2004). A system identification approach for video-based face recognition. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 175–178 Vol.4.

Alexandrov, A. D. (1957). Über eine verallgemeinerung der riemannschen geometrie. *Schr. Forschungsinst. Math.*, pages 33–84.

Arsigny, A., Fillard, P., Pennec, X., and Ayache, N. (2006). Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magn. Reson. Med.*, 56(2):411–421.
Bandulasiri, A., Bhattacharya, R., and Patrangenaru, V. (2009). Nonparametric inference for extrinsic means on size-and-(reflection)-shape manifolds with applications in medical imaging. *Journal of Multivariate Analysis*, 100(9):1867 – 1882.

Bandulasiri, A. and Patrangenaru, V. (2005). Algorithms for nonparametric inference on shape manifolds. *Proc. of JSM 2005, MN*, pages 1617–1622.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B.*, 57(1):289–300.

Bhattacharya, A. (2008). Statistical analysis on manifolds: a nonparametric approach for inference on shape spaces. *Sankhya*, 70:1–43.

Bhattacharya, A. and Bhattacharya, R. (2008). Statistics on Riemannian manifolds: asymptotic distribution and curvature. *Proceedings of the American Mathematical Society*, 136:2957–2967.

Bhattacharya, A. and Bhattacharya, R. (2012). *Nonparametric Inference on Manifolds: With Applications to Shape Spaces*. IMS Monograph #2. Cambridge University Press.

Bhattacharya, A. and Dunson, D. (2010). Nonparametric Bayesian density estimation on manifolds with applications to planar shapes. *Biometrika*, 97:851–865.

Bhattacharya, R. (2007). On the uniqueness of intrinsic mean. *Unpublished manuscript*.

Bhattacharya, R. and Lin, L. (2013). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *ArXiv e-prints*.

Bhattacharya, R. and Lin, L. (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *The Proceedings of the American Mathematical Society*, 145:413–428.

Bhattacharya, R. N., Lin, L., and Patrangenaru, V. (2016). *A Course in Mathematical Statistics and Large Sample Theory*. Springer Texts in Statistics, Springer.

Bhattacharya, R. N. and Patrangenaru, V. (2002). Nonparametric estimation of location and dispersion on riemannian manifolds. *J.Statist.Plan. Infer. Volume in honor of the 80th birthday of professor C.R.Rao*, 108:22–35.

Bhattacharya, R. N. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. *Ann. Statist.*, 31:1–29.

Bhattacharya, R. N. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds-ii. *Ann. Statist.*, 33:1225–1259.

Bookstein, F. L. (1991). *Morphometric Tools for Landmark data: Geometry and Biology*. Cambridge Univ. Press, Cambridge.

Boothby, W. (1986). *An Introduction to Differentiable Manifolds and Riemannian Geometry, 2nd ed.* Academic Press, New York.

Carlsson, G. (2009). Topology and Data. *Bulletin of the American Mathematical Society*, 46(2):255–308.

Chavel, I. (1984). *Eigenvalues in Riemannian Geometry*. Pure and Applied Mathematics. Elsevier Science.

Demmel, J., Eisenstat, S., Gilbert, J., Li, X., and Liu, J. (1999). A supernodal approach to sparse partial pivoting. *SIAM Journal on Matrix Analysis and Applications*, 20:720–755.

Dey, T. K. and Li, K. (2009). Persistence-based handle and tunnel loops computation revisited for speed up. *Computers & Graphics*, 33(3):351 – 358. {IEEE} International Conference on Shape Modelling and Applications 2009.

Dey, T. K., Li, K., Sun, J., and Cohen-Steiner, D. (2008). Computing geometry-aware handle and tunnel loops in 3d models. *ACM Trans. Graph.*, 27(3):45:1–45:9.

Do Carmo, M. (1992). *Riemannian Geometry*. Birkhäuser, Boston.
Dryden, I., Kume, A., Le, H., and Wood, A. T. (2008). A multi-dimensional scaling approach to shape analysis. *Biometrika*, 95(4):779–798.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. Wiley, New York.

Ellingson, L., Patrangenaru, V., and Ruymgaart, F. (2013). Nonparametric estimation of means on hilbert manifolds and extrinsic analysis of mean shapes of contours. *Journal of Multivariate Analysis*, 122:317–333.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1:209–230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.*, 2:615–629.

Fisher, N., Lewis, T., and Embleton, B. (1987). *Statistical Analysis of Spherical Data*. Cambridge Uni. Press, Cambridge.

Fisher, R. (1953). Dispersion on a sphere. *Proc. Roy. Soc. London Ser. A*, 217:295–305.

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

Gallot, S., Hulin, D., and Lafontaine, J. (1990). *Riemannian Geometry*. Universitext. Springer Verlag, Berlin.

Ghosh, J. and Ramamoorthi, R. (2003). *Bayesian Nonparametrics*. Springer Series in Statistics. Springer.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):285–339.

Goodlett, C., Davis, B., Jean, R., Gilmore, J., and Gerig, G. (2006). Improved correspondence for dti population studies via unbiased atlas building. *Med. Image Computing and Computer Assisted Intervention (MICCAI)*, 4191:260–267.

Gordon, C., Webb, D., and Wolpert, S. (1992). Isospectral plane domains and surfaces via riemannian orbifolds. *Inventiones mathematicae*, 110(1):1–22.

Gotsman, C., Gu, X., and Sheffer, A. (2003). Fundamentals of spherical parameterization for 3d meshes. *ACM Trans. Graph.*, 22(3):358–363.

Gromov, M. (1981). *Structures métriques pour les variétés riemanniennes*. Textes mathématiques. CEDIC/Fernand Nathan.

Hotelling, H. (1931). The generalization of student’s ratio. *Ann. Math. Statist.*, 2(3):360–378.

Hotz, T. and Huckemann, S. (2015). Intrinsic means on the circle: uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics*, 67(1):177–193.

Hotz, T., Skwerer, S., Huckemann, S., Le, H., Marron, J. S., Mattingly, J. C., Miller, E., Nolen, J., Owen, M., and Patrangenaru, V. (2013). Sticky central limit theorems on open books. *Annals of Applied Probability*, pages 2238–2258.

Huang, C., Styner, M., and Zhu, H. (2015). Penalized mixtures of offset-normal shape factor analyzers with application in clustering high-dimensional shape data. *J. Amer. Statist. Assoc.*, to appear.

Huckemann, S., Hotz, T., and Munk, A. (2010). Intrinsic shape analysis: geodesic pca for riemannian manifolds modulo isometric lie group actions (with discussions). *Statist. Sinica*, 20:1–100.

Irving, E. (1964). *Paleomagnetism and Its Application to Geological and Geographical Problems*. Wiley, New York.

Jain, V. and Zhang, H. (2007). A spectral approach to shape-based retrieval of articulated 3d models. *Comput. Aided Des.*, 39(5):398–407.
Jones, P. W., Maggioni, M., and Schul, R. (2008). Manifold parametrizations by eigenfunctions of the laplacian and heat kernels. *Proceedings of the National Academy of Sciences*, 105(6):1803–1808.

Kac, M. (1966). Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4):1–23.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30:509–554.

Kendall, D., Barden, D., Carne, T., and Le, H. (1999). *Shape and Shape Theory*. Wiley, New York.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. of the London Math. Soc.*, 16:81–121.

Kendall, D. G. (1989). A survey of the statistical theory of shape. *Statist. Sci.*, 4:87–120.

Kendall, W. (1990). Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proc. London Math. Soc.*, 61:371–406.

Kendall, W. and Le, H. (2011). Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables. *Brazilian Journal of Probability and Statistics*, 25(3):323–352.

Kindlmann, G., Tricoche, X., and Westin, C.-F. (2007). Delineating white matter structure in diffusion tensor MRI with anisotropy creases. *Medical Image Analysis*, 11 (5):492–502.

Le, H. (2001). Locating Fréchet means with application to shape spaces. *Adv. Appl.Prob.*, 33:324–338.

Lehmann, E. (1959). *Testing statistical hypotheses*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley.

Ma, Y., Soatto, S., Kosecká, J., and Sastry, S. (2005). *An Invitation to 3-D Vision: From Images to Geometric Models*. Interdisciplinary Applied Mathematics. Springer New York.

Mardia, K. and Jupp, P. (2000). *Directional Statistics*. Wiley, New York.

Mardia, K. and Patrangenaru, V. (2005). Directions and projective shapes. *Ann. Statist.*, 33:1666–1699.

McKilliam, R. G., Quinn, B. G., and Clarkson, I. V. L. (2012). Direction estimation by minimum squared arc length. *IEEE Transactions on Signal Processing*, 60(5):2115–2124.

Milnor, J. (1964). Eigenvalues of the laplace operator on certain manifolds. *Proceedings of the National Academy of Sciences*, 51(4):542.

Milnor, J. (1965). *Topology from the differentiable viewpoint*. University Press of Virginia.

Morr, J., Tu, Z., Apostolova, L., Green, A., Avedissian, C., Madsen, S., Parikshak, N., Toga, A., Jack Jr, C., Schuff, N., Weiner, M., and Thompson, P. (2000). Automated mappings of hippocampal atrophy in 1-year repeat mri data from 490 subjects with alzheimers disease, mild cognitive impairment, and elderly controls. *Neuroimage*, 45:S3–S15.

Patrangenaru, V., Liu, X., and Sugathadasa, S. (2010). A nonparametric approach to 3d shape analysis from digital camera images i. *Journal of Multivariate Analysis*, 101(1):11 – 31.

Pelletier, B. (2005). Kernel density estimation on riemannian manifolds. *Statistics and Probability Letters*, 73:297–304.

Ramsay, J. R. (2007). Current status of cognitive-behavioral therapy as a psychosocial treatment for adult attention-deficit/hyperactivity disorder. *Curr Psychiatry Rep.*, 9:273–281.
Reshetnyak, Y. (1968). Non-expansive maps in a space of curvature no greater than k. *Sibirsk. Mat. Zh.*, 9:918–927.

Reuter, M. (2006). *Laplace Spectra for Shape Recognition*. Books on Demand GmbH.

Reuter, M., Biasotti, S., Giorgi, D., Patan, G., and Spagnuolo, M. (2009). Discrete laplacebeltrami operators for shape analysis and segmentation. *Computers & Graphics, 33*(3):381 – 390. {IEEE} International Conference on Shape Modelling and Applications 2009.

Rosenberg, S. (1997). *The Laplacian on a Riemannian Manifold: An Introduction to Analysis on Manifolds*. EBSCO ebook academic collection. Cambridge University Press.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica*, 4:639–50.

Shamir, A. (2006). Segmentation and Shape Extraction of 3D Boundary Meshes. In Wyvill, B. and Wilkie, A., editors, *Eurographics 2006 - State of the Art Reports*. The Eurographics Association.

Sparr, G. (1992). Depth-computations from polihedral images. in g. sandimi, editor. *Proc. 2nd European Conf. on Computer Vision, pp 378-386. Springer Verlag*. Also in Image and Vision Computing, Vol 10, pp. 683-688.

Sturm, K. (2003). *Probability Measures on Metric Spaces of Nonpositive Curvature*. Preprint. SFB 611.

Sugathadasa, S. (2006). *Affine and Projective Shape Analysis with Applications*. Ph.D. dissertation, Texas Tech University.

Veeraraghavan, A., Roy-Chowdhury, A., and Chellappa, R. (2005). Matching shape sequences in video with applications in human movement analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(12):1896–1909.

Watson, G. S. (1983). *Statistics on Spheres*, volume 6. University Arkansas Lecture Notes in the Mathematical Sciences, Wiley, New York.

Weyl, H. (1911). Ueber die asymptotische verteilung der eigenwerte. *Nachrichten von der Gesellschaft der Wissenschaften zu Gittingen, Mathematisch-Physikalische Klasse*, pages 110–117.

Yuan, Y., Zhu, H., Lin, W., and Marron, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pages no–no.

Zelditch, S. (2000). Spectral determination of analytic bi-axisymmetric plane domains. *Geometric & Functional Analysis GAFA, 10*(3):628–677.

Ziezold, H. (1977). 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 Functions, Random Processes and of the Eighth European Meeting of Statisticians, A*:591–602. (Tech. Univ. Prague, Prague, 1974).