Sankhyā : The Indian Journal of Statistics
2008, Volume 70-A, Part 2, pp. 223-266

Statistical analysis on manifolds: A nonparametric approach for inference on shape spaces [1]

Abhishek Bhattacharya
Duke University, Durham, NC

Abstract

This article concerns nonparametric statistics on manifolds with special emphasis on *landmark based shape* spaces in which a k-ad, i.e., a set of k points or landmarks on an object or a scene, is observed in 2D or 3D, for purposes of identification, discrimination, or diagnostics. Two different notions of shape are considered: *reflection shape* invariant under all translations, scaling and orthogonal transformations, and *affine shape* invariant under all affine transformations. A computation of the *extrinsic mean* reflection shape, which has remained unresolved in earlier works, is given in arbitrary dimensions, enabling one to extend nonparametric inference on Kendall type shape manifolds from 2D to higher dimensions. For both reflection and affine shapes, two sample test statistics are constructed based on appropriate choice of orthonormal frames on the tangent bundle and computations of differentials of projection maps with respect to them at the sample extrinsic mean. The samples in consideration can be either independent or be the outcome of a matched pair experiment. Examples are included to illustrate the theory.

AMS (2000) *subject classification.* Primary 62H35; Secondary 62G20, 62F40.
Keywords and phrases. Manifold, equivariant embedding, shape space of k-ads, Fréchet function, extrinsic mean and variation, nonparametric analysis.

1 Introduction

Statistical analysis of a probability measure Q on a differentiable manifold M has diverse applications in directional and axial statistics, morphometrics, medical diagnostics and machine vision. In this article, we are concerned with the analysis of shapes of landmark based data, in which each observation consists of $k > m$ points in m-dimension called landmarks which represent k locations on an object. The configuration of k landmarks

[1]This research is partially supported by NSF Grant DMS 080611.

is called a k-ad. The choice of landmarks is generally made with expert help in the particular field of application. Depending on the way the data are collected or recorded, the appropriate shape of a k-ad is its orbit under a group of transformations. For example, one may look at k-ads modulo size and Euclidean rigid body motions of translation and rotation. The analysis of shapes under this invariance was pioneered by Kendall (1977, 1984) and Bookstein (1978). Bookstein's approach is primarily registration-based requiring two or three landmarks to be brought into a standard position by translation, rotation and scaling of the k-ad. For these shapes, we would prefer Kendall's more invariant view of a shape identified with the orbit under rotation (in m-dimension) of the k-ad centered at the origin and scaled to have unit size. The resulting shape spaces are called *similarity shape* spaces and denoted by Σ_m^k. A fairly comprehensive account of parametric inference on these spaces, with many references to the literature, may be found in Dryden and Mardia (1998).

Once we consider the orbits under all orthogonal transformations and scaling, we get the *reflection shape* spaces $R\Sigma_m^k$. It is possible to embed $R\Sigma_m^k$ into an Euclidean space and carry out *extrinsic analysis*. Such an embedding was first considered by Bandulasiri and Patrangenaru (2005) and later independently by Dryden et al. (2008). However the correct computation of the *extrinsic mean* reflection shape seems to have eluded earlier authors. It is computed in Corollary 6.3 in Section 6. This is one of the major results of this article.

Recently there has been much emphasis on the statistical analysis of other notions of shapes of k-ads, namely, *affine shapes* invariant under affine transformations, and *projective shapes* invariant under projective transformations. Reconstruction of a scene from two (or more) aerial photographs taken from a plane is one of the research problems in affine shape analysis. Potential applications of projective shape analysis include face recognition and robotics-for robots to visually recognize a scene (Mardia and Patrangenaru (2005)).

In this article, we will mainly focus on the reflection and affine shape spaces. We define the notions of extrinsic means and variations of probability distributions on general manifolds and compute them for the shape spaces. We develop nonparametric two sample tests to distinguish between two distributions by comparing the sample extrinsic means and variations.

The nonparametric methodology pursued here, along with the geometric and other mathematical issues that accompany it, stems from the earlier work of Bhattacharya and Patrangenaru(2002, 2003, 2005) and Bhattacharya and Bhattacharya (2008a, 2008b, 2008c). Examples of analysis with real data in Bhattacharya and Bhattacharya (2008a) suggest that appropriate nonparametric methods detect differences in shape distributions better than their parametric counterparts in the literature, for distributions that occur in applications.

The article is organized as follows. Section 2 introduces the notions of *Fréchet mean* and *variation* of a probability distribution Q on a metric space (M, ρ).

Section 3 outlines the theory of *extrinsic* analysis on general manifolds where ρ is the distance inherited by the manifold from an *embedding* J into some Euclidean space E^N. The image of M under J is a differentiable submanifold of E^N. A H-*equivariant* embedding (see Definition 3.1) with a relatively large group H preserves a corresponding group of symmetries of the manifold M and is therefore preferred in our analysis. The image of the Fréchet mean under the embedding J is the projection of the Euclidean mean of the push forward $Q \circ J^{-1}$ of Q on to $J(M)$ (see Proposition 3.1). This makes the Fréchet mean or *extrinsic mean* easy to compute in a number of important examples. In Section 3.1, we deduce the asymptotic distribution of the sample extrinsic mean. By the delta method, we linearly approximate the projection map by its differential into the tangent space of $J(M)$ at the embedding of the extrinsic mean of Q. With suitable choice of orthonormal basis for the tangent space, we derive coordinates for the difference between the embeddings of the sample extrinsic mean and the extrinsic mean of Q which have asymptotically Gaussian distribution. This is used to construct confidence region for the extrinsic mean of Q, both by an asymptotic chi-squared statistic and pivotal bootstrap methods. In Sectionn 3.2, we deduce the asymptotic distribution of the sample extrinsic variation and construct confidence intervals for the extrinsic variation of Q. The asymptotic theory is used in Section 3.3 to carry out two sample tests to identify differences between two probability distributions on the manifold. Appropriate tests are constructed for independent as well as matched pair samples. Matched pair samples arise, when, for example, we have two set of observations from the same subject (see Section 8.1). Hence the paired sample can be viewed as one sample in the product manifold $M \times M$. To do inference on the marginals of a probability distribution on $M \times M$, we apply the methods of Sections

3.1 and 3.2 to the tangent space of $J(M) \times J(M)$ at $(J(\mu_{1E}), J(\mu_{2E}))$ where μ_{iE}, $i = 1, 2$, are the extrinsic means of marginal distributions.

Section 4 provides a brief expository description of the geometries of the manifolds that arise in shape analysis.

Section 5 outlines the geometry of the planar similarity shape spaces and the general similarity shape spaces. The notions of a k-ad and how to represent its shape by an orbit under a group of transformations are introduced.

In Section 6, we derive an expression for the extrinsic mean on the reflection shape spaces under an equivariant embedding (see Corollary 6.3). We derive expressions for the tangent and normal spaces to the embedded submanifold through Proposition 6.1. We construct suitable orthonormal frames for the tangent space in Section 6.1 and use that to get asymptotic coordinates for the sample extrinsic mean. We require perturbation theory arguments for eigen values and eigen vectors to prove that the projection map of Section 3.1 is well defined and smooth. The methods of Section 3.3 are applied in Section 6.2 to carry out nonparametric inference on the reflection shape spaces.

In Section 7, analogous results are obtained for the affine shape spaces.

Finally, Section 8 illustrates the theory with two applications to real data. The data considered in Section 8.1 is a matched pair sample of 3D reflection shapes. We compare the extrinsic mean shapes and extrinsic variations in shape of the marginals by appropriate two sample tests. In the example in Section 8.2, we have an independent random sample of 2D affine shapes. After removing some outliers, we construct confidence regions for the mean shape and confidence intervals for the variation in shape.

When there are too many landmarks on the sample k-ads, making the dimension of the shape space close to the sample size, it becomes difficult to carry out inference on the mean shapes using pivotal bootstrap methods. That is because, in many simulations, the bootstrap covariance matrix is singular or close to being singular. Then one may compare only the first few principal scores of the coordinates of the sample extrinsic means, or use a nonpivotal bootstrap statistic, where one replaces the bootstrap covariance matrix by the sample covariance matrix. We try both these approaches for the examples in Section 8. Our analysis shows that the results obtained through appropriate bootstrap methods are consistent with those obtained by chi-squared approximation.

[Note: Henceforth, BP (...) stands for Bhattacharya and Patrangenaru (...) and BB (...) stands for Bhattacharya and Bhattacharya (...).]

2 Fréchet Mean and Variation on Metric Spaces

Let (M, ρ) be a metric space, ρ being a distance metrizing the topology of M. For a given probability distribution Q on (the Borel sigmafield of) M, define the *Fréchet function* of Q as

$$F(p) = \int_M \rho^2(p, x) Q(dx), \quad p \in M. \tag{2.1}$$

Now we define the Fréchet mean and Fréchet variation of Q. A general notion of a mean of a probability distribution on a metric space was first defined by Fréchet (1948). The concept of variation was introduced in BP (2002) where it has been referred to as the total variance.

DEFINITION 2.1. Suppose $F(p) < \infty$ for some $p \in M$. Then the set of all p for which $F(p)$ is the minimum value of F on M is called the *Fréchet mean set* of Q, denoted by C_Q. If this set is a singleton, say $\{\mu_F\}$, then μ_F is called the *Fréchet mean* of Q. The minimum value of F on M is called the *Fréchet variation* of Q and denoted by V. If $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) M-valued random variables defined on some probability space $(\Omega, \mathcal{F}, P)$ with common distribution Q, and $Q_n \doteq \frac{1}{n}\sum_{j=1}^n \delta_{X_j}$ is the corresponding empirical distribution, then the Fréchet mean set of Q_n is called the *sample Fréchet mean set*, denoted by C_{Q_n}. The Fréchet variation of Q_n is called the *sample Fréchet variation* and denoted by V_n.

Proposition 2.1 shows that under mild assumptions, the minimum value of F on M is attained, thereby proving that the Fréchet mean set is nonempty, as proved in Theorem 2.1, BP (2003).

PROPOSITION 2.1. *Suppose every closed and bounded subset of M is compact. If the Fréchet function F of Q is finite for some $p \in M$, then C_Q is nonempty and compact.*

Proposition 2.2 below proves the strong consistency of the *sample Fréchet mean* as an estimator of the Fréchet mean of Q. Here we define the sample

Fréchet mean as a measurable selection from the sample Fréchet mean set C_{Q_n}.

Proposition 2.2. *Suppose every closed and bounded subset of M is compact and the Fréchet function of Q is finite. Then given any $\epsilon > 0$, there exists an integer-valued random variable $N = N(\omega, \epsilon)$ and a P-null set $A(\omega, \epsilon)$ such that*

$$C_{Q_n} \subset C_Q^{\epsilon} \equiv \{p \in M : \rho(p, C_Q) < \epsilon\}, \ \forall n \geq N \tag{2.2}$$

outside of $A(\omega, \epsilon)$. In particular, if $C_Q = \{\mu_F\}$, then the sample Fréchet mean μ_{F_n} (any measurable selection from C_{Q_n}) is a strongly consistent estimator of μ_F.

Proof. See Theorem 2.3, BP (2003).

From Proposition 2.1 it follows that if the Fréchet function $F(p)$ is finite for some p, then the Fréchet variation V is finite and equals F(p) for all p in the Fréchet mean set C_Q. The sample Fréchet variation V_n is the value of F_n on the sample Fréchet mean set C_{Q_n}. Proposition 2.3 establishes the strong consistency of V_n as an estimator of V. For a proof, see BB (2008a).

Proposition 2.3. *Suppose every closed and bounded subset of M is compact, and F is finite on M. Then V_n is a strongly consistent estimator of V.*

Remark 2.1. It is known that a Riemannian manifold M which is complete (in its geodesic distance) satisfies the topological hypothesis of Propositions 2.1, 2.2 and 2.3: every closed and bounded subset of M is compact (see Theorem 2.8, Do Carmo (1992), pp. 146-147). The affine shape spaces considered in Section 7 are compact Riemannian manifolds and hence complete. When a manifold is not complete (similarity and reflection shape spaces, for example), we deduce other ways to establish consistency of the sample Fréchet mean and variation (e.g. Section 6, Remark 6.1).

Remark 2.2. Proposition 2.2 requires the Fréchet mean of Q to exist for the sample Fréchet mean to be a consistent estimator. However the sample Fréchet variation is a consistent estimator of the Fréchet variation of Q even when the Fréchet function F does not have a unique minimizer. We will investigate sufficient conditions for the existence of the Fréchet mean on shape spaces in the subsequent sections.

3 Extrinsic Analysis on Manifolds

From now on, we assume that M is a differentiable manifold of dimension d. To carry out nonparametric inference on M, one may use the Fréchet mean and variation to identify a probability distribution. Using the sample Fréchet mean and variation from a random sample on M, we can construct confidence regions for the population parameters, or given two such samples, we can distinguish between the underlying probability distributions by comparing the sample means and variations. The natural approach for nonparametric inference on a Riemannian manifold would be to use the geodesic distance in the definition of Fréchet function in (2.1) and derive expressions for Fréchet mean and variation. However it is simpler both mathematically and computationally to carry out an *extrinsic analysis* on M, by embedding it into some Euclidean space $E^N \approx \mathbb{R}^N$ via some map $J : M \to E^N$ such that both J and its derivative are injective, and for which $J(M)$ has the induced topology from E^N. Then J induces the metric

$$\rho(x, y) = \|J(x) - J(y)\| \tag{3.1}$$

on M, where $\|.\|$ denotes Euclidean norm ($\|u\|^2 = \sum_{i=1}^N {u_i}^2 \;\forall\, u = (u_1, u_2, \ldots, u_N)$). This is called the *extrinsic distance* on M. Among the possible embeddings, one seeks out *equivariant embeddings* which preserve many of the geometric features of M.

Definition 3.1. For a Lie group H acting on a manifold M, an embedding $J : M \to \mathbb{R}^N$ is *H-equivariant* if there exists a group homomorphism $\phi : H \to GL(N, \mathbb{R})$ such that

$$J(hp) = \phi(h)J(p) \;\forall p \in M,\; \forall h \in H. \tag{3.2}$$

Here $GL(N, \mathbb{R})$ is the *general linear group* of all $N \times N$ non-singular matrices.

For all our applications, H is compact.

In case $J(M) = \tilde{M}$ is a closed subset of E^N, for every $u \in E^N$ there exists a compact set of points in $\tilde{M}$ whose distance from u is the smallest among all points in $\tilde{M}$. We define this set to be the set of projections of u on $\tilde{M}$ and denote it by

$$P_{\tilde{M}}(u) = \{x \in \tilde{M} : \|x - u\| \leq \|y - u\| \;\forall y \in \tilde{M}\}. \tag{3.3}$$

If this set is a singleton, u is said to be a *nonfocal point* of E^N (w.r.t. $\tilde{M}$), otherwise it is said to be a *focal point* of E^N. Definition 3.2 below defines the extrinsic mean and variation of a probability distribution Q corresponding to the embedding J. The notion of extrinsic mean on a manifold was introduced independently by Hendricks and Landsman (1998) and Patrangenaru (1998), and later considered in detail in BP (2003, 2005).

DEFINITION 3.2. Let $(M, \rho), J$ be as above. Let Q be a probability measure on M such that the Fréchet function

$$F(x) = \int \rho^2(x, y) Q(dy) \tag{3.4}$$

is finite. The Fréchet mean set of Q is called the *extrinsic mean set* of Q and the Fréchet variation of Q is called the extrinsic variation of Q. If X_i, $i = 1, \ldots, n$ are iid observations from Q and $Q_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$, then the Fréchet mean set of Q_n is called the *sample extrinsic mean set* and the Fréchet variation of Q is called the *sample extrinsic variation.*

We define the *sample extrinsic mean* μ_{nE} to be a measurable selection from the sample extrinsic mean set. We say that Q has an extrinsic mean μ_E if the extrinsic mean set of Q is a singleton. Proposition 3.1 below gives a necessary and sufficient condition for that to hold.

PROPOSITION 3.1. *Let $\tilde{Q} = Q \circ J^{-1}$ be the image of Q in E^N. (a) If $\tilde{\mu} = \int_{E^N} u\tilde{Q}(du)$ is the mean of $\tilde{Q}$, then the extrinsic mean set of Q is given by $J^{-1}(P_{\tilde{M}}(\tilde{\mu}))$. (b) The extrinsic variation of Q equals*

$$V = \int_{E^N} \|x - \tilde{\mu}\|^2 \tilde{Q}(dx) + \|\tilde{\mu} - \mu\|^2 \tag{3.5}$$

where $\mu \in P_{\tilde{M}}(\tilde{\mu})$. (c) If $\tilde{\mu}$ is a nonfocal point of E^N, then the extrinsic mean of Q exists (as a unique minimizer of F).

PROOF. See Proposition 3.1, BP (2003).

3.1. Asymptotic distribution of the sample extrinsic mean. In this section, we assume that the extrinsic mean μ_E of Q is uniquely defined. Then the mean $\tilde{\mu}$ of $\tilde{Q}$ is a nonfocal point of E^N and hence the projection set in (3.3) is a well defined map in a neighborhood of $\tilde{\mu}$. Let us call that P ($P : E^N \to \tilde{M}$). Also in a neighborhood of a nonfocal point such as $\tilde{\mu}$, $P(.)$ is smooth. Let $\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$ be the sample mean of $Y_j = J(X_j)$, $j = 1, 2, \ldots, n$. Since $\bar{Y}$ converges to $\tilde{\mu}$ almost surely, for sample size large enough, $\bar{Y}$ is nonfocal and it can be shown that

$$\sqrt{n}[P(\bar{Y}) - P(\tilde{\mu})] = \sqrt{n}(d_{\tilde{\mu}}P)(\bar{Y} - \tilde{\mu}) + o_P(1) \tag{3.6}$$

where $d_{\tilde{\mu}}P$ is the differential (map) of the projection $P(.)$, which takes vectors in the tangent space of E^N at $\tilde{\mu}$ to tangent vectors of $\tilde{M}$ at $P(\tilde{\mu})$ (see BP(2005)). Since $\sqrt{n}(\bar{Y}-\tilde{\mu})$ has an asymptotic Gaussian distribution, and $d_{\tilde{\mu}}P$ is a linear map, from (3.6) it follows that $\sqrt{n}[P(\bar{Y})-P(\tilde{\mu})]$ has an asymptotic mean zero Gaussian distribution on the tangent space of $J(M)$ at $P(\tilde{\mu})$. Hence if we denote by T_j, the coordinates of $(d_{\tilde{\mu}}P)(Y_j-\tilde{\mu})$, $j=1,2,\ldots,n$ with respect to some orthonormal basis for $T_{P(\tilde{\mu})}\tilde{M}$, then

$$\sqrt{n}\bar{T} \xrightarrow{\mathcal{L}} N(0,\Sigma) \tag{3.7}$$

where Σ denotes the covariance matrix of T_1. Let $L_{\tilde{\mu}}: E^N \to T_{P(\tilde{\mu})}\tilde{M}$ denote the linear projection on to $T_{P(\tilde{\mu})}\tilde{M}$. Then from (3.6) and (3.7), it follows that

$$n(L_{\tilde{\mu}}[P(\bar{Y})-P(\tilde{\mu})])'\Sigma^{-1}L_{\tilde{\mu}}[P(\bar{Y})-P(\tilde{\mu})] \xrightarrow{\mathcal{L}} \mathcal{X}_d^2. \tag{3.8}$$

Using (3.8), we can construct the following confidence region for μ_E:

$$\{\mu_E = J^{-1}[P(\tilde{\mu})] : n(L[P(\tilde{\mu})-P(\bar{Y})])'\hat{\Sigma}^{-1}L[P(\tilde{\mu})-P(\bar{Y})] \le \mathcal{X}_d^2(1-\alpha)\} \tag{3.9}$$

with asymptotic confidence level $(1-\alpha)$. Here $L: E^N \to T_{P(\bar{Y})}\tilde{M}$ denotes the linear projection on to $T_{P(\bar{Y})}\tilde{M}$, $\hat{\Sigma}$ is the sample estimate of Σ and $\mathcal{X}_d^2(1-\alpha)$ is the $(1-\alpha)$-quantile of the chi-squared distribution with d degrees of freedom ($\mathcal{X}_d^2$). The corresponding pivotal bootstrap confidence region for μ_E is given by

$$\{\mu_E : n(L[P(\tilde{\mu})-P(\bar{Y})])'\hat{\Sigma}^{-1}L[P(\tilde{\mu})-P(\bar{Y})] \le c^*(1-\alpha)\}. \tag{3.10}$$

Here $c^*(1-\alpha)$ denotes the $(1-\alpha)$-quantile of the bootstrap distribution of

$$n(L^*[P(\bar{Y})-P(\bar{Y}^*)])'\Sigma^{*-1}L^*[P(\bar{Y})-P(\bar{Y}^*)],$$

Y_j^* is the bootstrap resample from Y_j, $j=1,\ldots,n$, $\bar{Y}^*$ and Σ^* are the bootsrap analogues of $\bar{Y}$ and $\hat{\Sigma}$ respectively, and L^* denotes the linear projection into $T_{P(\bar{Y}^*)}\tilde{M}$. In the example from Section 8, bootstrap methods yield much smaller confidence region for μ_E than by chi-squared approximation.

3.2. Asymptotic distribution of the sample extrinsic variation. Let V and V_n denote the extrinsic variation of Q and Q_n respectively. Let ρ be the extrinsic distance of (3.1).

PROPOSITION 3.2. *If Q has extrinsic mean μ_E and if* $\mathrm{E}\rho^4(X_1,\mu_E)<\infty$, *then*

$$\sqrt{n}(V_n - V) \xrightarrow{\mathcal{L}} N(0, \mathrm{Var}(\rho^2(X_1,\mu_E))) \tag{3.11}$$

Proof. From definition of V_n and V, it follows that

$$\begin{aligned} V_n - V &= \frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_{nE}) - \int_M \rho^2(x, \mu_E) Q(dx) \\ &= \frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_{nE}) - \frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_E) \\ &+ \frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_E) - \mathrm{E}\left[\rho^2(X_1, \mu_E)\right] \end{aligned} \tag{3.12}$$

where μ_{nE} is the sample extrinsic mean, i.e. some measurable selection from the sample extrinsic mean set. Note that

$$\begin{aligned} &\frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_{nE}) = \frac{1}{n}\sum_{j=1}^{n} \|Y_j - P(\bar{Y})\|^2 \\ =&\frac{1}{n}\sum_{j=1}^{n} \|Y_j - P(\tilde{\mu})\|^2 + \|P(\tilde{\mu}) - P(\bar{Y})\|^2 - 2\langle \bar{Y} - P(\tilde{\mu}), P(\bar{Y}) - P(\tilde{\mu})\rangle \end{aligned} \tag{3.13}$$

Substitute (3.13) in (3.12) to get that

$$\begin{aligned} V_n - V = \|P(\bar{Y}) - P(\tilde{\mu})\|^2 - 2\langle \bar{Y} - P(\tilde{\mu}), P(\bar{Y}) - P(\tilde{\mu})\rangle \\ + \frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_E) - \mathrm{E}\left[\rho^2(X_1, \mu_E)\right] \end{aligned} \tag{3.14}$$

which implies that

$$\sqrt{n}(V_n - V) = T_1 + T_2 \tag{3.15}$$

where

$$T_1 = \sqrt{n}\|P(\bar{Y}) - P(\tilde{\mu})\|^2 - 2\sqrt{n}\langle \bar{Y} - P(\tilde{\mu}), P(\bar{Y}) - P(\tilde{\mu})\rangle \text{ and} \tag{3.16}$$

$$T_2 = \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_E) - \mathrm{E}\left[\rho^2(X_1, \mu_E)\right]\right). \tag{3.17}$$

From the classical CLT, it follows that if $\mathrm{E}\rho^4(X_1, \mu_E) < \infty$, then

$$T_2 \xrightarrow{\mathcal{L}} N(0, \mathrm{Var}[\rho^2(X_1, \mu_E)]). \tag{3.18}$$

Compare the expression of T_1 with (3.6) to get that

$$T_1 = -2\langle d_{\tilde{\mu}} P(\bar{Y} - \tilde{\mu}), \tilde{\mu} - P(\tilde{\mu})\rangle + o_P(1). \tag{3.19}$$

From the definition of P, it follows that $P(\tilde{\mu}) = \text{argmin}_{p\in\tilde{M}} \|\tilde{\mu} - p\|^2$. Hence the Euclidean derivative of $\|\tilde{\mu} - p\|^2$ at $p = P(\tilde{\mu})$ must be orthogonal to $T_{P(\tilde{\mu})}\tilde{M}$, or

$$\tilde{\mu} - P(\tilde{\mu}) \in (T_{P(\tilde{\mu})}\tilde{M})^{\perp}.$$

Since $d_{\tilde{\mu}} P(\bar{Y} - \tilde{\mu}) \in T_{P(\tilde{\mu})}\tilde{M}$, the first term in the expression of T_1 in (3.19) is 0, and hence $T_1 = o_P(1)$. From (3.15) and (3.18), we conclude that

$$\sqrt{n}(V_n - V) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{\rho^2(X_j, \mu_E) - \mathrm{E}\left[\rho^2(X_1, \mu_E)\right]\right\} + o_P(1) \tag{3.20}$$

$$\xrightarrow{\mathcal{L}} N(0, \mathrm{Var}\left[\rho^2(X_1, \mu_E)\right]). \tag{3.21}$$

This completes the proof. □

REMARK 3.1. Although Proposition 2.3 does not require the uniqueness of the extrinsic mean of Q for V_n to be a consistent estimator of V, Proposition 3.2 breaks down in the case of non-uniqueness. (see BB (2008a)).

Using Proposition 3.2, we can construct the following confidence interval I for V:

$$I = \{V : \ V \in [V_n - \frac{s}{\sqrt{n}} Z(1 - \frac{\alpha}{2}), V_n + \frac{s}{\sqrt{n}} Z(1 - \frac{\alpha}{2})]\}. \tag{3.22}$$

The interval I has asymptotic confidence level of $(1 - \alpha)$. Here s^2 is the sample variance of $\rho^2(X_j, \mu_{nE})$, $j = 1, \ldots, n$ and $Z(1 - \frac{\alpha}{2})$ denotes the $(1 - \frac{\alpha}{2})$-quantile of $N(0, 1)$ distribution. From (3.22), we can also construct a pivotal bootstrap confidence interval for V, the details of which are left to the reader.

3.3. Two sample tests. In this section, we will use the asymptotic distribution of the sample extrinsic mean and variation to construct nonparametric tests to compare two probability distributions Q_1 and Q_2 on M.

3.3.1. Independent samples. Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two iid samples from Q_1 and Q_2 respectively that are mutually independent. Let μ_{iE} and V_i denote the extrinsic means and variations of Q_i, $i = 1, 2$ respectively. Similarly denote by $\hat{\mu}_{iE}$ and $\hat{V}_i$ the sample extrinsic means and variations. We want to test the hypothesis $H_0 : Q_1 = Q_2$.

We start by comparing the sample extrinsic means. Let $\tilde{X}_j = J(X_j)$, $\tilde{Y}_j = J(Y_j)$ be the embeddings of the sample points into E^N. Let μ_i be the mean of $\tilde{Q}_i = Q_i \circ J^{-1}$, $i = 1, 2$. Then under H_0, $\mu_1 = \mu_2 = \mu$ (say). Let $\hat{\mu}_i$, $i = 1, 2$ be the sample means of $\{\tilde{X}_j\}$ and $\{\tilde{Y}_j\}$ respectively. Then from (3.6), it follows that

$$\sqrt{n_i}[P(\hat{\mu}_i) - P(\mu)] = \sqrt{n_i}(d_\mu P)(\hat{\mu}_i - \mu) + o_P(1), \quad i = 1, 2. \tag{3.23}$$

Hence, if $n_i \to \infty$ such that $\frac{n_i}{n_1+n_2} \to p_i$, $0 < p_i < 1$, $p_1 + p_2 = 1$, then

$$\begin{aligned}\sqrt{n}(P(\hat{\mu}_1) - P(\hat{\mu}_2)) &= \sqrt{n} d_\mu P(\hat{\mu}_1 - \mu) - \sqrt{n} d_\mu P(\hat{\mu}_2 - \mu) + o_P(1) \\ &\xrightarrow{\mathcal{L}} N(0, \frac{\Sigma^1}{p_1} + \frac{\Sigma^2}{p_2}).\end{aligned} \tag{3.24}$$

Here $n = n_1 + n_2$ is the pooled sample size, Σ^i, $i = 1, 2$, are the covariance matrices of the coordinates of $d_\mu P(\tilde{X}_1 - \mu)$ and $d_\mu P(\tilde{Y}_1 - \mu)$ with respect to some chosen basis for $T_{P(\mu)}\tilde{M}$. We estimate μ by the pooled sample mean $\hat{\mu} = \frac{1}{n}(n_1\hat{\mu}_1 + n_2\hat{\mu}_2)$. Denote the coordinates of $\{d_{\hat{\mu}} P(\tilde{X}_j - \hat{\mu})\}_{j=1}^{n_1}$ and $\{d_{\hat{\mu}} P(\tilde{Y}_j - \hat{\mu})\}_{j=1}^{n_2}$ by $\{S_j^1\}_{j=1}^{n_1}$ and $\{S_j^2\}_{j=1}^{n_2}$ respectively. Let $\hat{\Sigma}^i$ $i = 1, 2$, denote the sample covariance matrices of $\{S_j^i\}_{j=1}^{n_i}$ $i = 1, 2$, respectively. Then if H_0 is true, the statistic

$$T_1 = (\bar{S^1} - \bar{S^2})' \left(\frac{1}{n_1}\hat{\Sigma}^1 + \frac{1}{n_2}\hat{\Sigma}^2\right)^{-1} (\bar{S^1} - \bar{S^2}) \tag{3.25}$$

converges in distribution to $\mathcal{X}^2$ distribution with d degrees of freedom, where d is the dimension of M. Hence we reject H_0 at asymptotic level α if $T_1 > \mathcal{X}_d^2(1 - \alpha)$.

Alternatively, under the null hypothesis $H_0 : \mu_{1E} = \mu_{2E}$, from (3.6), it follows that

$$\sqrt{n}[P(\hat{\mu}_1) - P(\hat{\mu}_2)] = \sqrt{n} d_{\mu_1} P(\hat{\mu}_1 - \mu_1) - \sqrt{n} d_{\mu_2} P(\hat{\mu}_2 - \mu_2) + o_P(1) \tag{3.26}$$

which implies that, for $\tilde{\mu} \in \tilde{M}$,

$$\begin{aligned}L_{\tilde{\mu}}[\sqrt{n}\{P(\hat{\mu}_1) - P(\hat{\mu}_2)\}] &= L_{1\tilde{\mu}}[\sqrt{n} d_{\mu_1} P(\hat{\mu}_1 - \mu_1)] \\ &\quad - L_{2\tilde{\mu}}[\sqrt{n} d_{\mu_2} P(\hat{\mu}_2 - \mu_2)] + o_P(1) \\ &\xrightarrow{\mathcal{L}} N(0, \frac{1}{p_1} L_{1\tilde{\mu}} \Sigma_1 L'_{1\tilde{\mu}} + \frac{1}{p_2} L_{2\tilde{\mu}} \Sigma_2 L'_{2\tilde{\mu}}).\end{aligned} \tag{3.27}$$

In (3.27), $L_{\tilde{\mu}}$ denotes the linear projection from E^N into $T_{\tilde{\mu}}\tilde{M}$ identified with $\mathbb{R}^d$. Similarly $L_{i\tilde{\mu}}$, $i = 1, 2$, denote the linear projections from $T_{P(\mu_i)}\tilde{M}$ into $T_{\tilde{\mu}}\tilde{M}$ or their associated matrices with respect to some chosen bases for the tangent spaces. For sake of simplicity we use the same notations. Finally Σ_i, $i = 1, 2$, are the covariance matrices of the coordinates of $d_{\mu_1}P(\tilde{X}_1 - \mu_1)$ and $d_{\mu_2}P(\tilde{Y}_1 - \mu_2)$ respectively. Note that $\tilde{\mu}$ can be any point on $\tilde{M}$ for (3.27) to hold. Using this one can construct the test statistic

$$T_2 = (L[P(\hat{\mu}_1) - P(\hat{\mu}_2)])' \left(\frac{1}{n_1}L_1\hat{\Sigma}_1L_1' + \frac{1}{n_2}L_2\hat{\Sigma}_2L_2'\right)^{-1} L[P(\hat{\mu}_1) - P(\hat{\mu}_2)] \tag{3.28}$$

to test if H_0 is true. In the statistic T_2, L is the linear projection from E^N into $T_p\tilde{M}$, where $p \in \tilde{M}$. L_i, $i = 1, 2$, are the matrices of linear projection from $T_{P(\hat{\mu}_i)}\tilde{M}$ into $T_p\tilde{M}$. The tangent spaces $T_p\tilde{M}$ and $T_{P(\hat{\mu}_i)}\tilde{M}$, $i = 1, 2$, are identified with $\mathbb{R}^d$ with respect to convenient basis frames. Again, p can be any point on $\tilde{M}$, but tangent space analysis is expected to provide better approximation to the asymptotic limit if we choose $p = P(\hat{\mu})$, $\hat{\mu}$ being the pooled sample mean. Finally $\hat{\Sigma}_i$, $i = 1, 2$, denote the sample covariance matrices of the coordinates of $\{d_{\hat{\mu}_1}P(\tilde{X}_j - \hat{\mu}_1)\}_{j=1}^{n_1}$ and $\{d_{\hat{\mu}_2}P(\tilde{Y}_j - \hat{\mu}_2)\}_{j=1}^{n_2}$ respectively with respect to the chosen basis for $T_{P(\hat{\mu}_1)}\tilde{M}$ and $T_{P(\hat{\mu}_2)}\tilde{M}$. Under H_0, $T_2 \xrightarrow{\mathcal{L}} \mathcal{X}_d^2$. Hence we reject H_0 at asymptotic level α if $T_2 > \mathcal{X}_d^2(1-\alpha)$.

When the sample sizes are not too large, it is more efficient to construct a bootstrap confidence region for $P(\mu_1) - P(\mu_2)$ using the test statistic T_2 in (3.28), and use that to test if H_0 is true. Let $\{X_j^*, j = 1, \ldots, n_1\}$ and $\{Y_j^*, j = 1, \ldots, n_2\}$ denote the bootstrap resamples from the original samples in $\tilde{M}$. Denote by μ_i^*, $i = 1, 2$, the bootstrap sample means, and by μ^*, the pooled sample mean. Let L_i^*, $i = 1, 2$, be the matrices of linear projections from $T_{P(\mu_i^*)}\tilde{M}$ into $T_{P(\mu^*)}\tilde{M}$, L^* be the linear projection from E^N into $T_{P(\mu^*)}\tilde{M}$, with the tangent spaces identified with $\mathbb{R}^d$, and Σ_i^*, $i = 1, 2$, be the bootstrap sample covariance matrices of the coordinates of $\{d_{\mu_1^*}P(X_j^* - \mu_1^*)\}$ and $\{d_{\mu_2^*}P(Y_j^* - \mu_2^*)\}$ respectively. Then the bootstrap version of T_2 in (3.28) is

$$\begin{aligned} &T_2^* = v^{*\prime}\Sigma^{*-1}v^* \text{ where} \\ &v^* = L^*[\{P(\mu_1^*) - P(\hat{\mu}_1)\} - \{P(\mu_2^*) - P(\hat{\mu}_2)\}] \text{ and} \\ &\Sigma^* = \frac{1}{n_1}L_1^*\hat{\Sigma}_1^*L_1^{*\prime} + \frac{1}{n_2}L_2^*\hat{\Sigma}_2^*L_2^{*\prime}. \end{aligned} \tag{3.29}$$

We reject H_0 at level α if $T_2 > c^*(1-\alpha)$, where $c^*(1-\alpha)$ is the $(1-\alpha)$-quantile of the bootstrap distribution of T_2^*.

Next we test if Q_1 and Q_2 have the same extrinsic variations, i.e. $H_0 : V_1 = V_2$. From Proposition 3.2 and the fact that the samples are independent, we get that, under H_0,

$$\sqrt{n}(\hat{V}_1 - \hat{V}_2) \xrightarrow{\mathcal{L}} N\left(0, \frac{\sigma_1^2}{p_1} + \frac{\sigma_2^2}{p_2}\right) \tag{3.30}$$

$$\Rightarrow \frac{\hat{V}_1 - \hat{V}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{3.31}$$

where $\sigma_1^2 = \mathrm{Var}[\rho^2(X_1, \mu_{1E})]$, $\sigma_2^2 = \mathrm{Var}[\rho^2(Y_1, \mu_{2E})]$ and s_1^2, s_2^2 are their sample estimates. Hence to test if H_0 is true, we can use the test statistic

$$T_3 = \frac{\hat{V}_1 - \hat{V}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}. \tag{3.32}$$

For a test of asymptotic size α, we reject H_0 if $|T_3| > Z(1 - \frac{\alpha}{2})$. We can also construct a bootstrap confidence interval for $V_1 - V_2$ and use that to test if $V_1 - V_2 = 0$. The details of that are left to the reader.

3.3.2. Matched paired samples. Next consider the case when (X_1, Y_1), …, (X_n, Y_n) is an iid sample from some distribution Q on $\bar{M} = M \times M$. Such samples arise, when for example two different treatments are applied to each subject in the sample (see Section 8.1).

Let X_j's have distribution Q_1 while Y_j's come from some distribution Q_2 on M. Our objective is to distinguish Q_1 and Q_2 by comparing the sample extrinsic means and variations. Since the X and Y samples are not independent, we cannot apply the methods of the earlier section. Instead we do our analysis on $\bar{M}$. Note that $\bar{M}$ is a differentiable manifold which can be embedded into $E^N \times E^N$ via the map

$$\bar{J} : \bar{M} \to E^N \times E^N, \ \bar{J}(x,y) = (J(x), J(y)).$$

Let $\tilde{Q} = Q \circ \bar{J}^{-1}$. Then if $\tilde{Q}_i$ has mean μ_i, $i = 1, 2$, then $\tilde{Q}$ has mean $\bar{\mu} = (\mu_1, \mu_2)$. The projection of $\bar{\mu}$ on to $\tilde{\bar{M}} \doteq \bar{J}(\bar{M})$ is given by $\bar{P}(\mu) =$

$(P(\mu_1), P(\mu_2))$. Hence if Q_i has extrinsic mean μ_{iE}, $i = 1, 2$, then Q has extrinsic mean $\bar{\mu}_E = (\mu_{1E}, \mu_{2E})$. Denote the paired sample as $Z_j \equiv (X_j, Y_j)$, $j = 1, \ldots, n$ and let $\hat{\bar{\mu}} = (\hat{\mu}_1, \hat{\mu}_2)$, $\hat{\bar{\mu}}_E = (\hat{\mu}_{1E}, \hat{\mu}_{2E})$ be the sample estimates of $\bar{\mu}$ and $\bar{\mu}_E$ respectively. From (3.6), it follows that

$$\sqrt{n}(\bar{P}(\hat{\bar{\mu}}) - \bar{P}(\bar{\mu})) = \sqrt{n} d_{\bar{\mu}} \bar{P}(\hat{\bar{\mu}} - \bar{\mu}) + o_P(1)$$

which can be written as

$$\sqrt{n} \begin{pmatrix} P(\hat{\mu}_1) - P(\mu_1) \\ P(\hat{\mu}_2) - P(\mu_2) \end{pmatrix} = \sqrt{n} \begin{pmatrix} d_{\mu_1} P(\hat{\mu}_1 - \mu_1) \\ d_{\mu_2} P(\hat{\mu}_2 - \mu_2) \end{pmatrix} + o_P(1) \tag{3.33}$$

Hence if H_0: $\mu_1 = \mu_2 = \mu$, then under H_0,

$$\begin{aligned} \sqrt{n} \begin{pmatrix} P(\hat{\mu}_1) - P(\mu) \\ P(\hat{\mu}_2) - P(\mu) \end{pmatrix} &= \sqrt{n} \begin{pmatrix} d_{\mu} P(\hat{\mu}_1 - \mu) \\ d_{\mu} P(\hat{\mu}_2 - \mu) \end{pmatrix} + o_P(1) \\ &\xrightarrow{\mathcal{L}} N(0, \Sigma = \begin{pmatrix} \Sigma^1 & \Sigma^{12} \\ \Sigma^{21} & \Sigma^2 \end{pmatrix}). \end{aligned} \tag{3.34}$$

In (3.34), Σ^i, $i = 1, 2$ are the same as in (3.24) and $\Sigma^{12} = (\Sigma^{21})'$ is the covariance between the coordinates of $d_\mu P(\tilde{X}_1 - \mu)$ and $d_\mu P(\tilde{Y}_1 - \mu)$. From (3.34), it follows that

$$\sqrt{n} d_\mu P(\hat{\mu}_1 - \hat{\mu}_2) \xrightarrow{\mathcal{L}} N(0, \Sigma^1 + \Sigma^2 - \Sigma^{12} - \Sigma^{21}). \tag{3.35}$$

This gives rise to the test statistic

$$T_{1p} = n(\bar{S^1} - \bar{S^2})'(\hat{\Sigma}^1 + \hat{\Sigma}^2 - \hat{\Sigma}^{12} - \hat{\Sigma}^{21})^{-1}(\bar{S^1} - \bar{S^2}) \tag{3.36}$$

where $\bar{S^1}$, $\bar{S^2}$, $\hat{\Sigma}^1$ and $\hat{\Sigma}^2$ are as in (3.25) and $\hat{\Sigma}^{12} = (\hat{\Sigma}^{21})'$ is the sample covariance between $\{S^1_j\}_{j=1}^n$ and $\{S^2_j\}_{j=1}^n$. If H_0 is true, T_{1p} converges in distribution to $\mathcal{X}^2_d$ distribution. Hence we reject H_0 at asymptotic level α if $T_{1p} > \mathcal{X}^2_d(1 - \alpha)$.

If the null hypothesis were $H_0 : \mu_{1E} = \mu_{2E}$, then from (3.33), it follows that under H_0,

$$\sqrt{n}[P(\hat{\mu}_1) - P(\hat{\mu}_2)] = \sqrt{n} d_{\mu_1} P(\hat{\mu}_1 - \mu_1) - \sqrt{n} d_{\mu_2} P(\hat{\mu}_2 - \mu_2) + o_P(1) \tag{3.37}$$

which implies that, for any $\tilde{\mu} \in \tilde{M}$,

$$\begin{aligned} & L_{\tilde{\mu}}[\sqrt{n}\{P(\hat{\mu}_1) - P(\hat{\mu}_2)\}] = L_{1\tilde{\mu}}[\sqrt{n} d_{\mu_1} P(\hat{\mu}_1 - \mu_1)] \\ & - L_{2\tilde{\mu}}[\sqrt{n} d_{\mu_2} P(\hat{\mu}_2 - \mu_2)] + o_P(1) \\ & \xrightarrow{\mathcal{L}} N(0, \Sigma) \text{ where} \\ & \Sigma = L_{1\tilde{\mu}} \Sigma_1 L'_{1\tilde{\mu}} + L_{2\tilde{\mu}} \Sigma_2 L'_{2\tilde{\mu}} - L_{1\tilde{\mu}} \Sigma_{12} L'_{2\tilde{\mu}} - L_{2\tilde{\mu}} \Sigma_{21} L'_{1\tilde{\mu}}. \end{aligned} \tag{3.38}$$

In (3.38), $L_{\tilde{\mu}}$, $L_{i\tilde{\mu}}$ and Σ_i, $i = 1, 2$ are the same as in (3.27), and $\Sigma_{12} = \Sigma_{21}'$ denotes the covariance between the coordinates of $d_{\mu_1} P(\tilde{X}_1 - \mu_1)$ and $d_{\mu_2} P(\tilde{Y}_1 - \mu_2)$. Hence to test if H_0 is true, one can use the test statistic

$$T_{2p} = nL[P(\hat{\mu}_1) - P(\hat{\mu}_2)]'\hat{\Sigma}^{-1}L[P(\hat{\mu}_1) - P(\hat{\mu}_2)] \text{ where} \tag{3.39}$$

$$\hat{\Sigma} = L_1\hat{\Sigma}_1 L_1' + L_2\hat{\Sigma}_2 L_2' - L_1\hat{\Sigma}_{12} L_2' - L_2\hat{\Sigma}_{21} L_1'. \tag{3.40}$$

In the statistic T_{2p}, L, L_i and $\hat{\Sigma}_i$, $i = 1, 2$ are as in (3.28) and $\hat{\Sigma}_{12} = (\hat{\Sigma}_{12})'$ denotes the sample covariance between the coordinates of $\{d_{\hat{\mu}_1} P(\tilde{X}_j - \hat{\mu}_1)\}_{j=1}^n$ and $\{d_{\hat{\mu}_2} P(\tilde{Y}_j - \hat{\mu}_2)\}_{j=1}^n$. Under H_0, $T_{2p} \xrightarrow{\mathcal{L}} \mathcal{X}_d^2$. Hence we reject H_0 at asymptotic level α if $T_{2p} > \mathcal{X}_d^2(1 - \alpha)$. In our example in Section 8.1, the two statistics T_{1p} and T_{2p} yield values which are quite close to each other.

One can also find a bootstrap confidence region for $P(\mu_1) - P(\mu_2)$ as in Section 3.3 and use that to test if H_0 is true. The details are left to the reader.

Let V_1 and V_2 denote the extrinsic variations of Q_1 and Q_2 and let $\hat{V}_1$, $\hat{V}_2$ be their sample analogues. Suppose we want to test the hypothesis, $H_0 : V_1 = V_2$. From (3.20), we get that

$$\begin{pmatrix} \sqrt{n}(\hat{V}_1 - V_1) \\ \sqrt{n}(\hat{V}_2 - V_2) \end{pmatrix} = \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{j=1}^n [\rho^2(X_j, \mu_{1E}) - \mathrm{E}\rho^2(X_1, \mu_{1E})] \\ \sum_{j=1}^n [\rho^2(Y_j, \mu_{2E}) - \mathrm{E}\rho^2(Y_1, \mu_{2E})] \end{pmatrix} + o_P(1) \tag{3.41}$$

$$\xrightarrow{\mathcal{L}} N\left(0, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right) \tag{3.42}$$

where $\sigma_{12} = \mathrm{Cov}(\rho^2(X_1, \mu_{1E}), \rho^2(Y_1, \mu_{2E}))$, σ_1^2 and σ_2^2 are as in (3.30). Hence if H_0 is true,

$$\sqrt{n}(\hat{V}_1 - \hat{V}_2) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}). \tag{3.43}$$

This gives rise to the test statistic,

$$T_{3p} = \frac{\sqrt{n}(\hat{V}_1 - \hat{V}_2)}{\sqrt{s_1^2 + s_2^2 - 2s_{12}}} \tag{3.44}$$

where s_1^2, s_2^2, s_{12} are sample estimates of $\sigma_1^2, \sigma_2^2, \sigma_{12}$ respectively. We reject H_0 at asymptotic level α if $|T_{3p}| > Z(1 - \frac{\alpha}{2})$. We can also get a $(1 - \alpha)$ level confidence interval for $V_1 - V_2$ using bootstrap resamples and use that to test if H_0 is true.

4 Geometry of Shape Manifolds

Many differentiable manifolds M naturally occur as submanifolds, or surfaces or hypersurfaces, of an Euclidean space. One example of this is the sphere $S^d = \{p \in \mathbb{R}^{d+1} \colon \|p\| = 1\}$. The shape spaces of interest here are not of this type. They are generally quotients of a Riemannian manifold N under the action of a transformation group G, i.e., $M = N/G$. A number of them are quotient spaces of $N = S^d$ under the action of a compact group G, i.e., the elements of the space are orbits in S^d traced out by the application of G. Among important examples of this kind are the Kendall's shape spaces and reflection shape spaces. In some cases the *action of the group is free*, i.e., $gp = p$ only holds for the identity element $g = e$. Then the elements of the orbit $O_p = \{gp \colon g \in G\}$ are in one-one correspondence with elements of G, and one can identify the orbit with the group. The orbit inherits the differential structure of the Lie group G. The tangent space T_pN at a point p may then be decomposed into a *vertical subspace* V_p of dimension that of the group G along the orbit space to which p belongs, and a *horizontal subspace* H_p which is orthogonal to it. The vertical subspace is isomorphic to the tangent space of G and the horizontal one can be identified with the tangent space of M at the orbit O_p. With this identification, M is a differentiable manifold of dimension that of N minus the dimension of G.

For carrying out an extrinsic analysis on M, we use a smooth map J from N into some Euclidean space E which is an embedding of M into that Euclidean space. Then the image $J(M)$ is a differentiable submanifold of E. If π denotes the projection map,

$$\pi : N \to M,\ \pi(p) = O_p,$$

then the tangent space of $J(M)$ at $J(\pi(p))$ is $dJ(H_p)$ where dJ denotes the differential of the map $J : N \to E$. Among all possible embeddings, we choose J to be equivariant under the action of a large group H on M. In most cases, H is compact.

5 Kendall's (Direct Similarity) Shape Spaces Σ_m^k

Kendall's shape spaces are quotient spaces S^d/G, under the action of the *special orthogonal group* $G = SO(m)$ of $m \times m$ orthogonal matrices with determinant $+1$. Important cases include $m = 2, 3$.

For the case $m = 2$, consider the space of all planar k-ads $(z_1, z_2, \ldots, z_k)$ $(z_j = (x_j, y_j))$, $k > 2$, excluding those with k identical points. The set of all centered and normed k-ads, say $u = (u_1, u_2, \ldots, u_k)$ comprise a unit sphere in a $(2k-2)$-dimensional vector space and is, therefore, a $(2k-3)$-dimensional sphere S^{2k-3}, called the *preshape sphere.* The group $G = SO(2)$ acts on the sphere by rotating each landmark by the same angle. The orbit under G of a point u in the preshape sphere can thus be seen to be the circle S^1, so that Kendall's planar shape space Σ_2^k can be viewed as the quotient space $S^{2k-3}/G \sim S^{2k-3}/S^1$, a $(2k-4)$-dimensional compact manifold. An algebraically simpler representation of Σ_2^k is given by the complex projective space $\mathbb{C}P^{k-2}$. For nonparametric extrinsic analysis on Σ_2^k, see BP (2003, 2005), BB (2008a). For many applications in archeology, astronomy, morphometrics, medical diagnosis, etc., see Bookstein (1986, 1997), Kendall (1989), Dryden and Mardia (1998), BP (2003, 2005), BB (2008a, 2008b, 2008c) and Small (1996).

When $m > 2$, consider a set of k points in $\mathbb{R}^m$, not all points being the same. Such a set is called a k-ad or a configuration of k landmarks. We will denote a k-ad by the $m \times k$ matrix, $x = (x_1, \ldots, x_k)$ where x_i, $i = 1, \ldots, k$ are the k landmarks from the object of interest. Assume $k > m$. The *direct similarity shape* of this k-ad is what remains after we remove the effects of translation, rotation and scaling. To remove translation, we substract the mean $\bar{x} = \frac{1}{k}\sum_{i=1}^{k} x_i$ from each landmark to get the centered k-ad $u = (x_1 - \bar{x}, \ldots, x_k - \bar{x})$. We remove the effect of scaling by dividing u by its euclidean norm to get

$$z = (\frac{x_1 - \bar{x}}{\|u\|}, \ldots, \frac{x_k - \bar{x}}{\|u\|}) = (z_1, z_2, \ldots, z_k). \tag{5.1}$$

This z is called the *preshape* of the k-ad x and it lies in the unit sphere S_m^k in the hyperplane $H_m^k = \{z \in \mathbb{R}^{m\times k} \colon z\mathbf{1}_k = 0\}$. Hence

$$S_m^k = \{z \in \mathbb{R}^{m\times k} \colon \mathrm{Trace(zz')} = 1,\ \mathrm{z}\mathbf{1}_\mathrm{k} = 0\} \tag{5.2}$$

Here $\mathbf{1}_k$ denotes the $k \times 1$ vector of all ones. Thus the *preshape sphere* S_m^k may be identified with the sphere S^{km-m-1}. Then the shape of the k-ad x is the orbit of z under left multiplication by $m \times m$ rotation matrices. In other words $\Sigma_m^k = S_m^k/SO(m)$. One can also remove the effect of translation from the original k-ad x by postmultiplying the centered k-ad u by a *Helmert* matrix H which is a $k \times (k-1)$ matrix satisfying $H'H = I_{k-1}$ and $\mathbf{1}_k' H = 0$.

The resulting k-ad $\tilde{u} = uH$ lies in $\mathbb{R}^{m\times(k-1)}$. Then the preshape of x is $\tilde{z} = \tilde{u}/\|\tilde{u}\|$ and the preshape sphere can be represented as

$$S_m^k = \{z \in \mathbb{R}^{m\times(k-1)} \colon \text{Trace}(\text{zz}') = 1\} \tag{5.3}$$

The advantage of using this representation of S_m^k is that there is no linear constraint on the coordinates of z and hence analysis becomes simpler. However, now the choice of the preshape depends on the choice of H which can vary. In most cases, including applications, we will represent the preshape of x as in (5.1) and the preshape sphere by (5.2).

For $m > 2$, the direct similarity shape space Σ_m^k fails to be a manifold. That is because the action of $SO(m)$ is not in general free. Indeed, the orbits of preshapes under $SO(m)$ have different dimensions in different regions (see, e.g., Kendall et al. (1999) and Small (1996)). To avoid that, one may consider the shape of only those k-ads whose preshapes have rank at least $m-1$. This subset is a manifold but is not complete (in its geodesic distance).

6 Reflection (Similarity) Shape Spaces $R\Sigma_m^k$

Consider now the *reflection shape* of a k-ad as defined in Section 5, but with $SO(m)$ replaced by the larger *orthogonal group* $O(m)$ of all $m \times m$ orthogonal matrices (with determinants either +1 or -1). Then the reflection (similarity) shape of a k-ad x is given by the orbit

$$\sigma(x) = \sigma(z) = \{Az \colon A \in O(m)\} \tag{6.1}$$

where z is the preshape of x in S_m^k. For the action of $O(m)$ on S_m^k to be free and the reflection shape space to be a Riemannian manifold, we consider only those shapes where the columns of z span $\mathbb{R}^m$. The set of all such z is called the nonsingular part of S_m^k and denoted by NS_m^k. Then the *reflection (similarity) shape space* is defined as

$$R\Sigma_m^k = \{\sigma(z) \colon z \in S_m^k,\ \text{rank}(z) = m\} = NS_m^k/O(m) \tag{6.2}$$

which is a Riemannian manifold of dimension $km - m - 1 - m(m-1)/2$.

It has been shown that the map

$$J : R\Sigma_m^k \to S(k, \mathbb{R}),\ J(\sigma(z)) = z'z \tag{6.3}$$

is an embedding of the reflection shape space into $S(k,\mathbb{R})$ (see Bandulasiri and Patrangenaru (2005), Bandulasiri et al. (2007) and Dryden et al. (2008)). It is H-equivariant where $H = O(k)$ acts on the right: $A\sigma(z) \doteq \sigma(zA')$, $A \in O(k)$. Indeed, then $J(A\sigma(z)) = \phi(A)J(\sigma(z))$ where

$$\phi : O(k) \to GL(k,\mathbb{R}),\ \phi(A)\colon S(k,\mathbb{R}) \to S(k,\mathbb{R}),\ \phi(A)B = ABA'.$$

It is easy to show that $\phi(A)$ is an isometry and ϕ is a group homomorphism. Define M_m^k as the set of all $k \times k$ positive semi-definite matrices of rank m and trace 1. Then the image of $R\Sigma_m^k$ under the embedding J in (6.3) is

$$J(R\Sigma_m^k) = \{A \in M_m^k\colon A\mathbf{1}_k = 0\}. \tag{6.4}$$

If we represent the preshape sphere S_m^k as in (5.3), then $M_m^k = J(R\Sigma_m^{k+1})$. Hence M_m^k is a submanifold (not complete) of $S(k,\mathbb{R})$ of dimension $km - 1 - m(m-1)/2$.

The main results of this section are Theorems 6.1, 6.2 and their corollaries. Together with Corollary 6.1, Theorem 6.1 derives the extrinsic mean of a probability distribution on the reflection shape space under the embedding (6.3). A recent computation of this given in Dryden et al. (2008) is incorrect (See Remark 6.2). Proposition 6.1 below identifies the tangent and normal spaces of M_m^k. The expression for the tangent space has also been derived in Dryden et al. (2008). The derivation of the tangent space in the proof of Proposition 6.1 is different from the one in there and is included here for the sake of readability.

Proposition 6.1. *Let $A \in M_m^k$. (a) The tangent space of M_m^k at A is given by*

$$T_A M_m^k = \{U \begin{pmatrix} T & S \\ S' & 0 \end{pmatrix} U'\colon T \in S(m,\mathbb{R}),\ \mathrm{trace}(T) = 0\} \tag{6.5}$$

where $A = UDU'$ is a singular value decomposition (s.v.d.) of A, $U \in SO(m)$ and $D = \mathrm{diag}(\lambda_1,\ldots,\lambda_k)$, $\lambda_1 \geq \ldots \geq \lambda_k \geq 0$. (b) The orthocomplement of the tangent space in $S(k,\mathbb{R})$ or the normal space is given by

$$(T_A M_m^k)^\perp = \{U \begin{pmatrix} \lambda I_m & 0 \\ 0 & T \end{pmatrix} U'\colon \lambda \in \mathbb{R},\ T \in S(k-m,\mathbb{R})\} \tag{6.6}$$

Proof. Represent the preshape of a $(k+1)-$ad x by the $m \times k$ matrix z where $\|z\|^2 = \mathrm{Trace}(zz') = 1$ and let S_m^{k+1} be the preshape sphere,

$$S_m^{k+1} = \{z \in \mathbb{R}^{m\times k}\colon \|z\| = 1\}.$$

Let NS_m^{k+1} be the nonsingular part of S_m^{k+1}, i.e.,

$$NS_m^{k+1} = \{z \in S_m^{k+1} \colon \text{rank}(z) = m\}.$$

Then $R\Sigma_m^{k+1} = NS_m^{k+1}/O(m)$ and $M_m^k = J(R\Sigma_m^{k+1})$. The map

$$J : R\Sigma_m^{k+1} \longrightarrow S(k, \mathbb{R}),\ J(\sigma(z)) = z'z = A$$

is an embedding. Hence

$$T_A M_m^k = dJ_{\sigma(z)}(T_{\sigma(z)} R\Sigma_m^{k+1}). \tag{6.7}$$

Note that $T_{\sigma(z)} R\Sigma_m^{k+1}$ can be identified with the horizontal subspace H_z of $T_z S_m^{k+1}$ which is

$$H_z = \{v \in R^{m\times k} \colon\ \text{trace}(zv') = 0,\ zv' = vz'\} \tag{6.8}$$

(see Kendall et. al. (1999)). Consider the map

$$\tilde{J} : NS_m^{k+1} \to S(k, \mathbb{R}),\ \tilde{J}(z) = z'z. \tag{6.9}$$

Its derivative is a isomorphism between the horizontal subspace of $TNS_m^{k+1} \equiv TS_m^{k+1}$ and TM_m^k. The derivative is given by

$$d\tilde{J} : TS_m^{k+1} \to S(k, \mathbb{R}),\ d\tilde{J}_z(v) = z'v + v'z. \tag{6.10}$$

Hence

$$T_A M_m^k = d\tilde{J}_z(H_z) = \{z'v + v'z \colon v \in H_z\}. \tag{6.11}$$

From the description of H_z in (6.8) and using the fact that z has full row rank, it follows that

$$H_z = \{zv \colon v \in \mathbb{R}^{k\times k},\ \text{trace}(z'zv) = 0, zvz' \in S(m, \mathbb{R})\}. \tag{6.12}$$

From (6.11) and (6.12), we get that

$$T_A M_m^k = \{Av + v'A \colon AvA \in S(k, \mathbb{R}),\ \text{trace}(Av) = 0\}. \tag{6.13}$$

Let $A = UDU'$ be a s.v.d. of A as in the statement of the proposition. Using the fact that A has rank m, (6.13) can be written as

$$\begin{aligned} T_A M_m^k &= \{U(Dv + v'D)U' : DvD \in S(k, \mathbb{R}),\ \text{trace}(Dv) = 0\} \\ &= \{U \begin{pmatrix} T & S \\ S' & 0 \end{pmatrix} U' \colon T \in S(m, \mathbb{R}),\ \text{Trace}(T) = 0\}. \end{aligned} \tag{6.14}$$

This proves part (a). From the definition of orthocomplement and (6.14), we get that

$$\begin{aligned}(T_A M_m^k)^\perp &= \{v \in S(k,\mathbb{R})\colon \operatorname{trace}(v'w) = 0 \ \forall \ w \in T_A M_m^k\} \\ &= \{U \begin{pmatrix} \lambda I_m & 0 \\ 0 & R \end{pmatrix} U'\colon \lambda \in \mathbb{R},\ R \in S(k-m,\mathbb{R})\} \qquad (6.15)\end{aligned}$$

where I_m is the $m \times m$ identity matrix. This proves (b) and completes the proof. □

For a $k \times k$ positive semi definite matrix μ with rank at least m, its projection into M_m^k is defined as

$$P(\mu) = \{A \in M_m^k\colon \|\mu - A\|^2 = \underset{x \in M_m^k}{\operatorname{argmin}} \|\mu - x\|^2\} \qquad (6.16)$$

if this set is non empty. The following theorem shows that the projection set is nonempty and derives formula for the projection matrices.

Theorem 6.1. *$P(\mu)$ is non empty and consists of*

$$A = \sum_{j=1}^{m} (\lambda_j - \bar{\lambda} + \frac{1}{m}) U_j U_j' \qquad (6.17)$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ are the ordered eigen values of μ; $U_1, U_2, \ldots, U_k$ are some corresponding orthonormal eigen vectors and $\bar{\lambda} = \frac{1}{m}\sum_{j=1}^m \lambda_j$.

Proof. Let

$$f(x) = \|\mu - x\|^2,\ x \in S(k,\mathbb{R}). \qquad (6.18)$$

If f has a minimizer A in M_m^k then $(\operatorname{grad} f)(A) \in T_A(M_m^k)^\perp$ where grad denotes the Euclidean derivative operator. But $(\operatorname{grad} f)(A) = 2(A - \mu)$. Hence if A minimizes f, then

$$A - \mu = U^A \begin{pmatrix} \lambda I_m & 0 \\ 0 & T \end{pmatrix} U^{A\prime} \qquad (6.19)$$

where $U^A = (U_1^A, U_2^A, \ldots, U_k^A)$ is a $k \times k$ matrix consisting of an orthonormal basis of eigen vectors of A corresponding to its ordered eigen values $\lambda_1^A \geq \lambda_2^A \geq \ldots \geq \lambda_m^A > 0 = \ldots = 0$. From (6.19) it follows that

$$\mu U_j^A = (\lambda_j^A - \lambda) U_j^A;\ j = 1, 2, \ldots, m. \qquad (6.20)$$

Hence $\{\lambda_j^A - \lambda\}_{j=1}^m$ are eigen values of μ with $\{U_j^A\}_{j=1}^m$ as corresponding eigen vectors. Since these eigen values are ordered, this implies that there exists a singular value decomposition of μ: $\mu = \sum_{j=1}^{k} \lambda_j U_j U_j'$, and a set of indices $S = \{i_1, i_2, \ldots, i_m\}$, $1 \leq i_1 < i_2 < \ldots < i_m \leq k$ such that

$$\lambda_j^A - \lambda = \lambda_{i_j} \text{ and} \tag{6.21}$$

$$U_j^A = U_{i_j},\ j = 1, \ldots, m. \tag{6.22}$$

Add the equations in (6.21) to get $\lambda = \frac{1}{m} - \bar{\lambda}$ where $\bar{\lambda} = \frac{\sum_{j \in S} \lambda_j}{m}$. Hence

$$A = \sum_{j \in S} (\lambda_j - \bar{\lambda} + \frac{1}{m}) U_j U_j'. \tag{6.23}$$

Since $\sum_{j=1}^{k} \lambda_j = 1$, therefore $\bar{\lambda} \leq 1/m$ and $\lambda_j - \bar{\lambda} + \frac{1}{m} > 0\ \forall j \in S$. Hence A is positive semi definite of rank m. It is easy to check that trace(A)=1 and hence $A \in M_m^k$.

It can be shown that among the matrices A of the form (6.23), the function f defined in (6.18) is minimized when

$$S = \{1, 2, \ldots, m\}. \tag{6.24}$$

Define $M_{\leq m}^k$ as the set of all $k \times k$ positive semi-definite matrices of rank $\leq m$ and trace $= 1$. This is a compact subset of $S(k, \mathbb{R})$. Hence f restricted to $M_{\leq m}^k$ attains a minimum value. Let A_0 be a corresponding minimizer. If rank(A_0) $< m$, say $= m_1$, then A_0 minimizes f restricted to $M_{m_1}^k$. $M_{m_1}^k$ is a Riemannian manifold (it is $J(R\Sigma_{m_1}^{k+1})$). Hence A_0 must have the form

$$A_0 = \sum_{j=1}^{m_1} (\lambda_j - \bar{\lambda} + \frac{1}{m_1}) U_j U_j' \tag{6.25}$$

where $\bar{\lambda} = \frac{\sum_{j=1}^{m_1} \lambda_j}{m_1}$. But if one defines

$$A = \sum_{j=1}^{m} (\lambda_j - \bar{\lambda} + \frac{1}{m}) U_j U_j' \tag{6.26}$$

with $\bar{\lambda} = \frac{\sum_{j=1}^{m} \lambda_j}{m}$, then it is easy to check that $f(A) < f(A_0)$. Hence A_0 cannot be a minimizer of f over $M_{\leq m}^k$, that is, a minimizer must have rank $= m$. Then it lies in M_m^k and from (6.23) and (6.24), it follows that it has the form as in (6.26). This completes the proof. □

Let Q be a probability distribution on $R\Sigma_m^k$ and let $\tilde{\mu}$ be the mean of $\tilde{Q} \equiv Q \circ J^{-1}$ in $S(k, \mathbb{R})$. Then $\tilde{\mu}$ is positive semi definite of rank at least m and $\tilde{\mu}\mathbf{1}_k = 0$. Theorem 6.1 can be used to get the formula for the extrinsic mean set of Q. This is obtained in Corollary 6.1.

Corollary 6.1. *(a) The projection of $\tilde{\mu}$ into $J(R\Sigma_m^k)$ is given by*

$$P_{J(R\Sigma_m^k)}(\tilde{\mu}) = \{A\colon A = \sum_{j=1}^{m}(\lambda_j - \bar{\lambda} + \frac{1}{m})U_j {U_j}'\} \tag{6.27}$$

where $\lambda_1 \geq \ldots \geq \lambda_k$ are the ordered eigen values of $\tilde{\mu}$, $U_1, \ldots, U_k$ are corresponding orthonormal eigen vectors and $\bar{\lambda} = \frac{\sum_{j=1}^{m}\lambda_j}{m}$. (b) The projection set in (6.27) *is a singleton and Q has a unique extrinsic mean μ_E iff $\lambda_m > \lambda_{m+1}$. Then $\mu_E = \sigma(F)$ where $F = (F_1, \ldots, F_m)'$, $F_j = \sqrt{\lambda_j - \bar{\lambda} + \frac{1}{m}}U_j$.*

Proof. Since $\tilde{\mu}\mathbf{1}_k = 0$, therefore $U_j'\mathbf{1}_k = 0\ \forall j \leq m$. Hence any A in (6.27) lies in $J(R\Sigma_m^k)$ Now part (a) follows from Theorem 6.1 using the fact that $J(R\Sigma_m^k) \subseteq M_m^k$. For simplicity, let us denote $\lambda_j - \bar{\lambda} + \frac{1}{m}$, $j = 1, \ldots, m$ by λ_j^*. To prove part (b), note that if $\lambda_m = \lambda_{m+1}$, clearly $A_1 = \sum_{j=1}^{m}\lambda_j^* U_j {U_j}'$ and $A_2 = \sum_{j=1}^{m-1}\lambda_j^* U_j {U_j}' + \lambda_m^* U_{m+1}U_{m+1}'$ are two distinct elements in the projection set of (6.27). Consider next the case $\lambda_m > \lambda_{m+1}$. Let $\tilde{\mu} = U\Lambda U' = V\Lambda V'$ be two different s.v.d. of $\tilde{\mu}$. Then $U'V$ consists of orthonormal eigen vectors of $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$. The fact $\lambda_m > \lambda_{m+1}$ implies that

$$U'V = \begin{pmatrix} V_{11} & 0 \\ 0 & V_{22} \end{pmatrix} \tag{6.28}$$

where $V_{11} \in SO(m)$ and $V_{22} \in SO(k-m)$. Write

$$\Lambda = \begin{pmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{pmatrix}.$$

Then $\Lambda U'V = U'V\Lambda$ implies $\Lambda_{11}V_{11} = V_{11}\Lambda_{11}$ and $\Lambda_{22}V_{22} = V_{22}\Lambda_{22}$. Hence

$$\begin{aligned}
&\sum_{j=1}^{m}\lambda_j^* V_j V_j' \\
=&U\sum_{j=1}^{m}\begin{pmatrix} \lambda_j^*(V_{11})_j(V_{11})_j' & 0 \\ 0 & 0 \end{pmatrix}U' \\
=&U\begin{pmatrix} \Lambda_{11} + (\frac{1}{m} - \bar{\lambda})I_m & 0 \\ 0 & 0 \end{pmatrix}U' \\
=&\sum_{j=1}^{m}\lambda_j^* U_j U_j'.
\end{aligned}$$

This proves that the projection set in (6.27) is a singleton when $\lambda_m > \lambda_{m+1}$. Then for any F in part (b) and A as in (6.27), $A = F'F = J(\sigma(F))$. This proves part (b) and completes the proof. □

From Proposition 3.1 and Corollary 6.1, it follows that the extrinsic variation of Q has the following expression:

$$V = \int_{J(R\Sigma_m^k)} \|x - \tilde{\mu}\|^2 \tilde{Q}(dx) + \|\tilde{\mu} - A\|^2, \;\; A \in P_{J(R\Sigma_m^k)}(\tilde{\mu}).$$

$$= \int_{J(R\Sigma_m^k)} \|x\|^2 \tilde{Q}(dx) + m(\frac{1}{m} - \bar{\lambda})^2 - \sum_{j=1}^{m} \lambda_j^2. \tag{6.29}$$

REMARK 6.1. From the proof of Theorem 6.1 and Corollary 6.1, it follows that the extrinsic mean set C_Q of Q is also the extrinsic mean set of Q restricted to $M_{\le m}^k$. Since $M_{\le m}^k$ is a compact metric space, from Proposition 2.1, it follows that C_Q is compact. Let $X_1, X_2, \ldots, X_n$ be an iid sample from Q and let μ_{nE} and V_n be the sample extrinsic mean and variation respectively. Then from Proposition 2.3, it follows that V_n is a consistent estimator of V. From Proposition 2.2, it follows that if Q has a unique extrinsic mean μ_E, then μ_{nE} is a consistent estimator of μ_E.

REMARK 6.2. In Dryden et al. (2008), the mean ϕ-shape is defined as

$$\phi(\tilde{\mu}) = \frac{\sum_{j=1}^{m} \lambda_j U_j U_j'}{\sum_{j=1}^{m} \lambda_j}.$$

The article states that this is the 'natural projection' of $\tilde{\mu}$ on to $J(R\Sigma_m^k)$. However one can easily check that the distance between $\phi(\tilde{\mu})$ and $\tilde{\mu}$ is in general greater than the distance of the latter from the correct expression of the projection $P_{J(R\Sigma_m^k)}(\tilde{\mu})$ given in Corollary 6.1. The article also states that, "The mean ϕ-shape so defined is the 'extrinsic' mean reflection shape ... in the sense of Bhattacharya & Patrangenaru (2003, 2005) and Hendriks & Landsman (1998)" which is clearly incorrect.

6.1. Asymptotic distribution of the sample mean reflection shape. Let $X_1, \ldots, X_n$ be an iid sample from some probability distribution Q on $R\Sigma_m^k$ and let μ_{nE} be the sample extrinsic mean (any measurable selection from the sample extrinsic mean set). In the last section, we saw that if Q has a unique extrinsic mean μ_E, that is if the mean $\tilde{\mu}$ of $\tilde{Q} = Q \circ J^{-1}$ is a nonfocal point of $S(k, \mathbb{R})$, then μ_{nE} converges a.s. to μ_E as $n \to \infty$. Also from the calculations of Section 3.1, it follows that if the projection map $P \equiv P_{J(R\Sigma_m^k)}$ is continuously differentiable at $\tilde{\mu}$, then $\sqrt{n}[J(\mu_{nE}) - J(\mu_E)]$ has asymptotic mean zero Gaussian distribution on $T_{J(\mu_E)}J(R\Sigma_m^k)$. To find the asymptotic coordinates and the asymptotic dispersion matrix, we need to compute the differential of P at $\tilde{\mu}$ (if it exists).

Consider first the map $P : N(\tilde{\mu}) \to S(k, \mathbb{R})$, $P(\mu) = \sum_{j=1}^{m}(\lambda_j(\mu) - \bar{\lambda}(\mu) + 1/m)U_j(\mu)U_j(\mu)'$ as in Theorem 6.1. Here $N(\tilde{\mu})$ is an open neighborhood of $\tilde{\mu}$ in $S(k, \mathbb{R})$ where P is defined. Hence for $\mu \in N(\tilde{\mu})$, $\lambda_m(\mu) > \lambda_{m+1}(\mu)$. It can be shown that P is smooth on $N(\tilde{\mu})$ (see Theorem 6.2). Let $\gamma(t) = \tilde{\mu} + tv$ be a curve in $N(\tilde{\mu})$ with $\gamma(0) = \tilde{\mu}$ and $\dot{\gamma}(0) = v \in S(k, \mathbb{R})$. Let $\tilde{\mu} = U\Lambda U'$, $U = (U_1, \ldots, U_k)$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k)$ be a s.v.d. of $\tilde{\mu}$ as in Corollary 6.1. Then

$$\gamma(t) = U(\Lambda + tU'vU)U' = U\tilde{\gamma}(t)U' \tag{6.30}$$

where $\tilde{\gamma}(t) = \Lambda + tU'vU$. Then $\tilde{\gamma}(t)$ is a curve in $S(k, \mathbb{R})$ starting at Λ. Say $\tilde{v} = \dot{\tilde{\gamma}}(0) = U'vU$. From (6.30) and the definition of P, we get that

$$P[\gamma(t)] = UP[\tilde{\gamma}(t)]U'. \tag{6.31}$$

Differentiate (6.31) at $t = 0$, and noting that $\frac{\mathrm{d}}{\mathrm{d}t}P[\gamma(t)]|_{t=0} = d_{\tilde{\mu}}P(v)$ and $\frac{\mathrm{d}}{\mathrm{d}t}P[\tilde{\gamma}(t)]|_{t=0} = d_{\Lambda}P(\tilde{v})$, to get that

$$d_{\tilde{\mu}}P(v) = Ud_{\Lambda}P(\tilde{v})U'. \tag{6.32}$$

Let us find $\frac{\mathrm{d}}{\mathrm{d}t}P[\tilde{\gamma}(t)]|_{t=0}$. For that without loss of generality, we may assume that $\lambda_1 > \lambda_2 > \ldots > \lambda_k$. That is because, the set of all such matrices forms an open dense set of $S(k, \mathbb{R})$. Then we can choose a s.v.d. for $\tilde{\gamma}(t)$: $\tilde{\gamma}(t) = \sum_{j=1}^{k}\lambda_j(t)e_j(t)e_j(t)'$ such that $\{e_j(t), \lambda_j(t)\}_{j=1}^{k}$ are some smooth functions of t satisfying $e_j(0) = e_j$ and $\lambda_j(0) = \lambda_j$, where $\{e_j\}_{j=1}^{k}$ is the canonical basis for $\mathbb{R}^k$. Since $e_j(t)'e_j(t) = 1$, we get by differentiating,

$$e_j'\dot{e}_j(0) = 0, \ j = 1, \ldots, k. \tag{6.33}$$

Also since $\tilde{\gamma}(t)e_j(t) = \lambda_j(t)e_j(t)$, we get that

$$\tilde{v}e_j + \Lambda\dot{e}_j(0) = \lambda_j\dot{e}_j(0) + \dot{\lambda}_j(0)e_j, \ j = 1, \ldots, k. \tag{6.34}$$

Consider the orthonormal basis (frame) for $S(k, \mathbb{R})$: $\{E_{ab} : 1 \le a \le b \le k\}$ defined as

$$E_{ab} = \begin{cases} \frac{1}{\sqrt{2}}(e_ae_b^t + e_be_a^t) & \text{if } a < b \\ e_ae_a^t & \text{if } a = b. \end{cases} \tag{6.35}$$

Let $\tilde{v} = E_{ab}$, $1 \leq a \leq b \leq k$. From equations (6.33) and (6.34), we get that

$$\dot{e}_j(0) = \begin{cases} 0 & \text{if } a = b \text{ or } j \notin \{a, b\} \\ 2^{-1/2}(\lambda_a - \lambda_b)^{-1} e_b & \text{if } j = a < b \\ 2^{-1/2}(\lambda_b - \lambda_a)^{-1} e_a & \text{if } j = b > a \end{cases} \tag{6.36}$$

and

$$\dot{\lambda}_j(0) = \begin{cases} 1 & \text{if } j = a = b \\ 0 & \text{otherwise} \end{cases} \tag{6.37}$$

Since

$$P[\tilde{\gamma}(t)] = \sum_{j=1}^{m} [\lambda_j(t) - \bar{\lambda}(t) + \frac{1}{m}] e_j(t) e_j(t)'$$

where $\bar{\lambda}(t) = \frac{1}{m} \sum_{j=1}^{m} \lambda_j(t)$, therefore

$$\dot{\bar{\lambda}}(0) = \frac{1}{m} \sum_{j=1}^{m} \dot{\lambda}_j(0),$$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} P[\tilde{\gamma}(t)]|_{t=0} &= \sum_{j=1}^{m} [\dot{\lambda}_j(0) - \dot{\bar{\lambda}}(0)] e_j e_j' \\ &\quad + \sum_{j=1}^{m} [\lambda_j - \bar{\lambda} + \frac{1}{m}][e_j \dot{e}_j(0)' + \dot{e}_j(0) e_j']. \end{aligned} \tag{6.38}$$

Take $\dot{\tilde{\gamma}}(0) = \tilde{v} = E_{ab}$, $1 \leq a \leq b \leq k$ in (6.38). From equations (6.36) and (6.37), we get that

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t} P[\tilde{\gamma}(t)]|_{t=0} &= d_\Lambda P(E_{ab}) \\ &= \begin{cases} E_{ab} & \text{if } a < b \leq m, \\ E_{aa} - \frac{1}{m} \sum_{j=1}^{m} E_{jj} & \text{if } a = b \leq m, \\ (\lambda_a - \bar{\lambda} + \frac{1}{m})(\lambda_a - \lambda_b)^{-1} E_{ab} & \text{if } a \leq m < b \leq k, \\ 0 & \text{if } m < a \leq b \leq k. \end{cases} \end{aligned} \tag{6.39}$$

Then from (6.32) and (6.39), we get that

$$d_{\tilde{\mu}} P(U E_{ab} U') = \begin{cases} U E_{ab} U' & \text{if } a < b \leq m, \\ U \left(E_{aa} - \frac{1}{m} \sum_{j=1}^{m} E_{jj} \right) U' & \text{if } a = b \leq m, \\ (\lambda_a - \bar{\lambda} + \frac{1}{m})(\lambda_a - \lambda_b)^{-1} U E_{ab} U' & \text{if } a \leq m < b \leq k, \\ 0 & \text{if } m < a \leq b \leq k. \end{cases} \tag{6.40}$$

From the description of the tangent space $T_{P(\tilde{\mu})}M_m^k$ in (6.5), it is clear that

$$d_{\tilde{\mu}}P(UE_{ab}U') \in T_{P(\tilde{\mu})}M_m^k \ \forall \ a \leq b.$$

Let us denote by

$$F_{ab} = UE_{ab}U', \ 1 \leq a \leq m, a < b \leq k, \tag{6.41}$$

$$F_a = UE_{aa}U', \ 1 \leq a \leq m. \tag{6.42}$$

Then from (6.40), we get that

$$d_{\tilde{\mu}}P(UE_{ab}U') = \begin{cases} F_{ab} & \text{if } 1 \leq a < b \leq m, \\ F_a - \bar{F} & \text{if } a = b \leq m, \\ \left(\lambda_a - \bar{\lambda} + \frac{1}{m}\right)(\lambda_a - \lambda_b)^{-1} F_{ab} & \text{if } 1 \leq a \leq m < b \leq k, \\ 0 & \text{otherwise} \end{cases} \tag{6.43}$$

where $\bar{F} = \frac{1}{m}\sum_{a=1}^m F_a$. Note that the vectors $\{F_{ab}, F_a\}$ in (6.41) and (6.42) are orthonormal and $\sum_{a=1}^m (F_a - \bar{F}) = 0$. Hence from (6.43), we conclude that the subspace spanned by $d_{\tilde{\mu}}P(UE_{ab}U')$ has dimension

$$\frac{m(m-1)}{2} + m - 1 + m(k-m)$$
$$= km - m - \frac{m(m-1)}{2} = \dim(M_m^k).$$

This proves that

$$T_{P(\tilde{\mu})}M_m^k = \text{Span}\{d_{\tilde{\mu}}P(UE_{ab}U')\}_{a \leq b}.$$

Consider the orthonormal basis $\{UE_{ab}U' : 1 \leq a \leq b \leq k\}$ of $S(k, \mathbb{R})$. Define

$$\tilde{F}_a = \sum_{j=1}^m H_{aj}F_j, \ 1 \leq a \leq m-1 \tag{6.44}$$

where H is a $(m-1) \times m$ Helmert matrix, that is $HH' = I_{m-1}$ and $H\mathbf{1}_m = 0$. Then the vectors $\{F_{ab}\}$ defined in (6.41) and $\{\tilde{F}_a\}$ defined in (6.44) together form an orthonormal basis of $T_{P(\tilde{\mu})}M_m^k$. This is proved in Theorem 6.2.

Theorem 6.2. *Let $\tilde{\mu}$ be a nonfocal point in $S(k, \mathbb{R})$. Let $\tilde{\mu} = U\Lambda U'$ be a s.v.d. of $\tilde{\mu}$. (a) The projection map $P : N(\tilde{\mu}) \to S(k, \mathbb{R})$ is smooth and its derivative $dP : S(k, \mathbb{R}) \to TM_m^k$ is given by* (6.40). *(b) The vectors (matrices) $\{F_{ab} : 1 \leq a \leq m, a < b \leq k\}$ defined in* (6.41) *and $\{\tilde{F}_a : 1 \leq a \leq$

$(m-1)\}$ defined in (6.44) *together form an orthonormal basis of $T_{P(\tilde{\mu})}M^k_m$.*
(c) Let $A \in S(k,\mathbb{R}) \equiv T_{\tilde{\mu}}S(k,\mathbb{R})$ have coordinates $((a_{ij}))_{1\le i\le j\le k}$ with respect to the orthonormal basis $\{UE_{ij}U'\}$ of $S(k,\mathbb{R})$. That is,

$$A = \sum\sum_{1\le i\le j\le k} a_{ij}UE_{ij}U',$$

$$a_{ij} = \langle A, UE_{ij}U'\rangle = \begin{cases} \sqrt{2}U_i'AU_j & \text{if } i<j \\ U_i'AU_i & \text{if } i=j. \end{cases}$$

Then $d_{\tilde{\mu}}P(A)$ has coordinates

$$\begin{aligned} &a_{ij},\ 1\le i<j\le m,\\ &\tilde{a}_i,\ 1\le i\le (m-1),\\ &\left(\lambda_i-\bar{\lambda}+\frac{1}{m}\right)(\lambda_i-\lambda_j)^{-1}a_{ij},\ 1\le i\le m<j\le k \end{aligned}$$

w.r.t. the orthonormal basis $\{F_{ij}: 1\le i<j\le m\}$, $\{\tilde{F}_i : 1\le i\le (m-1)\}$ and $\{F_{ij}: 1\le i\le m<j\le k\}$ of $T_{P(\tilde{\mu})}M^k_m$. Here

$$\begin{aligned} &\mathbf{a}\equiv (a_{11},a_{22},\dots,a_{mm})',\\ &\tilde{\mathbf{a}}\equiv (\tilde{a}_1,\tilde{a}_2,\dots,\tilde{a}_{m-1})' = H\mathbf{a}. \end{aligned}$$

Proof. Let $\mu \in N(\tilde{\mu})$ have ordered eigen values $\lambda_1(\mu)\ge\lambda_2(\mu)\ge\dots\ge\lambda_k(\mu)$ with corresponding orthonormal eigen vectors $U_1(\mu),U_2(\mu),\dots,U_k(\mu)$. Then from Perturbation theory, it follows that if $\lambda_m(\mu)>\lambda_{m+1}(\mu)$, then

$$\mu\mapsto \text{Span}(U_1(\mu),\dots,U_m(\mu)),\ \sum_{i=1}^m\lambda_i(\mu)$$

are smooth maps into their respective co domains (see Dunford and Schwartz (1958), p. 598). Write

$$P(\mu)=\sum_{j=1}^m\lambda_j(\mu)U_j(\mu)U_j(\mu)' + \left(\frac{1}{m}-\bar{\lambda}(\mu)\right)\sum_{j=1}^m U_j(\mu)U_j(\mu)'.$$

Then $\sum_{j=1}^m U_j(\mu)U_j(\mu)'$ is the projection matrix of the subspace $\text{Span}(U_1(\mu), \dots, U_m(\mu))$, which is a smooth function of μ. Also $\sum_{j=1}^m \lambda_jU_j(\mu)U_j(\mu)'$ is the projection of μ into the subspace $\text{Span}(U_1(\mu)U_1(\mu)',\dots,U_m(\mu)U_m(\mu)')$ and hence is a smooth function of μ. Thus $\mu\mapsto P(\mu)$ is a smooth map on $N(\tilde{\mu})$.

This proves part (a).
From (6.43), we conclude that $\{F_{ab} : 1 \leq a \leq m, a < b \leq k\}$ and $\{F_a - \bar{F} : 1 \leq a \leq m\}$ span $T_{P(\tilde{\mu})}M_m^k$. It is easy to check from the definition of H that $\text{Span}\{\tilde{F}_a : 1 \leq a \leq (m-1)\} = \text{Span}\{F_a - \bar{F} : 1 \leq a \leq m\}$. Also since $\{F_a\}$ are mutually orthogonal, so are $\{\tilde{F}_a\}$. This proves that $\{F_{ab} : 1 \leq a \leq m, a < b \leq k\}$ and $\{\tilde{F}_a : 1 \leq a \leq (m-1)\}$ together form an orthonormal basis of $T_{P(\tilde{\mu})}M_m^k$, which is claimed in part (b). If $A = \sum\sum_{1\leq i\leq j\leq k} a_{ij}UE_{ij}U'$, then

$$\begin{aligned} d_{\tilde{\mu}}P(A) &= \sum\sum a_{ij} d_{\tilde{\mu}}P(UE_{ij}U') \\ &= \sum\sum_{1\leq i<j\leq m} a_{ij}F_{ij} + \sum_{i=1}^{m} a_{ii}(F_i - \bar{F}) \\ &+ \sum_{i=1}^{m}\sum_{j=m+1}^{k} a_{ij}(\lambda_i - \bar{\lambda} + \frac{1}{m})(\lambda_i - \lambda_j)^{-1}F_{ij} \qquad (6.45) \\ &= \sum\sum_{1\leq i<j\leq m} a_{ij}F_{ij} + \sum_{i=1}^{m-1} \tilde{a}_i\tilde{F}_i \\ &+ \sum_{i=1}^{m}\sum_{j=m+1}^{k} (\lambda_i - \bar{\lambda} + \frac{1}{m})(\lambda_i - \lambda_j)^{-1} a_{ij}F_{ij}. \qquad (6.46) \end{aligned}$$

This proves part (c). To get (6.46) from (6.45), we use the fact that $\sum_{i=1}^{m} a_{ii}(F_i - \bar{F}) = \sum_{i=1}^{m-1} \tilde{a}_i\tilde{F}_i$. To show that, denote by F, the matrix $(F_1, \ldots, F_m)$, by $F-\bar{F}$ the matrix $(F_1-\bar{F}, \ldots, F_m-\bar{F})$ and by $\tilde{F}$ the matrix $(\tilde{F}_1, \ldots, \tilde{F}_{m-1})$. Then

$$\begin{aligned} &\sum_{i=1}^{m-1} \tilde{a}_i\tilde{F}_i = \tilde{F}\tilde{\mathbf{a}} \\ =&\tilde{F}H\mathbf{a} = F(I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m')\mathbf{a} \\ =&(F - \bar{F})\mathbf{a} = \sum_{i=1}^{m} a_{ii}(F_i - \bar{F}). \end{aligned}$$

This completes the proof.

Corollary 6.2. *Consider the projection map restricted to* $S_0(k,\mathbb{R}) \equiv \{A \in S(k,\mathbb{R})\colon A\mathbf{1}_k = 0\}$. *Then its derivative is given by*

$$\begin{aligned} dP\colon S_0(k,\mathbb{R}) &\to TJ(R\Sigma_m^k), \\ d_{\tilde\mu}P(A) &= \sum\sum_{1\le i<j\le m} a_{ij}F_{ij} + \sum_{i=1}^{m-1} \tilde a_i \tilde F_i \\ &\quad + \sum_{i=1}^{m}\sum_{j=m+1}^{k-1} (\lambda_i - \bar\lambda + \frac{1}{m})(\lambda_i-\lambda_j)^{-1} a_{ij}F_{ij}. \end{aligned} \tag{6.47}$$

Hence $d_{\tilde\mu}P(A)$ *has coordinates* $\{a_{ij}, 1\le i<j\le m\}$, $\{\tilde a_i, 1\le i\le (m-1)\}$, $\{(\lambda_i-\bar\lambda+\frac{1}{m})(\lambda_i-\lambda_j)^{-1}a_{ij}, 1\le i\le m<j<k\}$ *w.r.t. the orthonormal basis* $\{F_{ij}: 1\le i<j\le m\}$, $\{\tilde F_i: 1\le i\le (m-1)\}$ *and* $\{F_{ij}: 1\le i\le m<j<k\}$ *of* $T_{P(\tilde\mu)}J(R\Sigma_m^k)$.

Proof. Follows from the fact that

$$T_{P(\tilde\mu)}J(R\Sigma_m^k) = \{v \in T_{P(\tilde\mu)}M_m^k\colon v\mathbf{1}_k = 0\}$$

and $\{F_{ij}\colon j=k\}$ lie in $T_{P(\tilde\mu)}J(R\Sigma_m^k)^\perp$. □

Consider the same set up as in Section 3.1. Let $Y_j = J(X_j)$, $j=1,\dots,n$ be the image of the sample into $J(R\Sigma_m^k)$. Let d be the dimension of $R\Sigma_m^k$. Let T_j, $j=1,\dots,n$ be the coordinates of $d_{\tilde\mu}P(Y_j-\tilde\mu)$ in $T_{P(\tilde\mu)}J(R\Sigma_m^k)\approx\mathbb{R}^d$. Then from (3.6), it follows that

$$\sqrt{n}[P(\bar Y) - P(\tilde\mu)] = \sqrt{n}\bar T + o_P(1) \xrightarrow{d} N(0, \mathrm{Cov}(T_1)).$$

We can get expression for T_j and hence $\bar T$ and $\mathrm{Cov}(T_1)$ from Corollary 6.2 as follows. Define

$$\begin{aligned} (Y_j)_{ab} &= \begin{cases} \sqrt{2}U_a'Y_jU_b \text{ if } 1\le a<b\le k, \\ U_a'Y_jU_a - \lambda_a \text{ if } a=b, \end{cases} \\ S_j &= H((Y_j)_{11}, (Y_j)_{22}, \dots, (Y_j)_{mm})', \\ (T_j)_{ab} &= \begin{cases} (Y_j)_{ab} \text{ if } 1\le a<b\le m, \\ (S_j)_a \text{ if } 1\le a=b\le (m-1), \\ (\lambda_a-\bar\lambda+\frac{1}{m})(\lambda_a-\lambda_b)^{-1}(Y_j)_{ab} \text{ if } 1\le a\le m<b<k. \end{cases} \end{aligned} \tag{6.48}$$

Then $T_j \equiv ((T_j)_{ab})$ is the vector of asymptotic coordinates of Y_j in $T_{P(\tilde\mu)}J(R\Sigma_m^k)\approx\mathbb{R}^d$. Now we can construct confidence regions for the extrinsic mean shape μ_E as in Section 3.1.

6.2. Two sample tests on $R\Sigma_m^k$. Now we are in the same set up as in Section 3: there are two samples on $R\Sigma_m^k$ and we want to test if they come from the same distribution, by comparing their sample extrinsic means and variations. To use the test statistic T_1 in (3.25) to compare the extrinsic means, we need to find the coordinates of $\{d_{\hat{\mu}}P(\tilde{X}_j - \hat{\mu})\}$ and $\{d_{\hat{\mu}}P(\tilde{Y}_j - \hat{\mu})\}$ in $T_{P(\hat{\mu})}J(R\Sigma_m^k)$. We get those from Corollary 6.2 as in (6.48). To use the test statistic T_2 in (3.28), we need expressions for $L : S(k,\mathbb{R}) \to T_{P(\hat{\mu})}J(R\Sigma_m^k)$ and $L_i : T_{P(\hat{\mu}_i)}J(R\Sigma_m^k) \to T_{P(\hat{\mu})}J(R\Sigma_m^k)$, $i = 1, 2$. Let $\hat{\mu} = U\Lambda U'$ be a s.v.d. of $\hat{\mu}$. Consider the orthonormal basis $\{UE_{ij}U' : 1 \le i \le j \le k\}$ of $S(k,\mathbb{R})$ and the orthonormal basis of $T_{P(\hat{\mu})}J(R\Sigma_m^k)$ derived in Corollary 6.2. Then if $A \in S(k,\mathbb{R})$ has coordinates $\{a_{ij} : 1 \le i \le j \le k\}$, it is easy to show that $L(A)$ has coordinates $\{a_{ij} : 1 \le i < j \le m\}$, $\{\tilde{a}_i : 1 \le i \le m-1\}$ and $\{a_{ij} : 1 \le i \le m < j < k\}$ in $T_{P(\hat{\mu})}J(R\Sigma_m^k)$. If we label the bases of $T_{P(\hat{\mu}_i)}J(R\Sigma_m^k)$ as $\{v_1^i, \ldots, v_d^i\}$, $i = 1, 2$ and that of $T_{P(\hat{\mu})}J(R\Sigma_m^k)$ as $\{v_1, \ldots, v_d\}$, then one can show that L_i is the $d \times d$ matrix with coordinates

$$(L_i)_{ab} = \langle v_a, v_b^i \rangle \;\; 1 \le a, b \le d, \;\; i = 1, 2.$$

7 Affine Shape Spaces $A\Sigma_m^k$

The *affine shape* of a k-ad x with landmarks in $\mathbb{R}^m$ may be defined as the orbit of this k-ad under the group of all *affine transformations* $x \mapsto F(x) = Ax + b$, where A is an arbitrary $m \times m$ non-singular matrix and b is an arbitrary point in $\mathbb{R}^m$. Note that two k-ads $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k)$, (x_j, $y_j \in \mathbb{R}^m$ for all j) have the same affine shape if and only if the centered k-ads $u = (u_1, u_2, \ldots, u_k) = (x_1 - \bar{x}, \ldots, x_k - \bar{x})$ and $v = (v_1, v_2, \ldots, v_k) = (y_1 - \bar{y}, \ldots, y_k - \bar{y})$ are related by a transformation $Au \doteq (Au_1, \ldots, Au_k) = v$. The centered k-ads lie in a linear subspace of $\mathbb{R}^m$ of dimension $m(k-1)$. Assume $k > m+1$. The *affine shape space* is then defined as the quotient space $H(m,k)/GL(m,\mathbb{R})$, where $H(m,k)$ consists of all centered k-ads whose landmarks span $\mathbb{R}^m$, and $GL(m,\mathbb{R})$ is the general linear group on $\mathbb{R}^m$ of all $m \times m$ nonsingular matrices. $H(m,k)$ can be identified with $\mathbb{R}^{m(k-1)}$. $GL(m,\mathbb{R})$ has the relative topology (and distance) of $\mathbb{R}^{m^2}$ and hence is a manifold of dimension m^2. Since the action of $GL(m,\mathbb{R})$ on $H(m,k)$ is free, it follows that $A\Sigma_m^k$ is a manifold of dimension $m(k-1) - m^2$. For $u, v \in H(m,k)$, since $Au = v$ iff $u'A' = v'$, and as A varies $u'A'$ generates the linear subspace L of $H(m,k)$ spanned by the m rows of u. The affine shape of u, (or of x), is identified with this subspace. Thus $A\Sigma_m^k$ may be identified with the set of all m dimensional subspaces of $\mathbb{R}^{k-1}$, namely, the *Grassmannian* $G_m(k-1)$-a result of Sparr (1992) (Also see Boothby (1986), pp. 63-64, 362-363). Affine shape spaces arise in certain problems of bioinformatics, cartography, machine vision and pattern recognition (see Mardia et al. (2005), Berthilsson and Heyden (1999), Berthilsson and Astrom (1999), Sepiashvili et al. (2003) and Sparr (1992)).

Let u be a centered k-ad in $H(m,k)$, and let $\sigma(u)$ denote its affine shape, which is the orbit

$$\sigma(u) = \{Au\colon A \in GL(m,\mathbb{R})\}.$$

Consider the map

$$J : A\Sigma_m^k \to S(k,\mathbb{R}),\ J(\sigma(u)) \equiv A = FF' \tag{7.1}$$

where $F = (f_1, f_2, \ldots, f_m)$ is an orthonormal basis for the row space of u. It has been shown that J is an embedding of $A\Sigma_m^k$ into $S(k,\mathbb{R})$, equivariant under the action of $O(k)$ (see Dimitric (1996)). In (7.1), A is the projection (matrix) onto the subspace spanned by the rows of u. Hence through the embedding J, we identify a m-dimensional subspace of $\mathbb{R}^{k-1}$ with the projection map (matrix) onto that subspace. Since A is a projection matrix, it is characterized by

$$A^2 = A,\ A = A' \text{ and } \mathrm{trace}(A) = \mathrm{rank}(A) = m.$$

Also since u is a centered k-ad, that is, the rows of u are orthogonal to $\mathbf{1}_k$, therefore $A\mathbf{1}_k = 0$. Hence the image of $A\Sigma_m^k$ into $S(k,\mathbb{R})$ under the embedding J is given by

$$J(A\Sigma_m^k) = \{A \in S(k,\mathbb{R})\colon A^2 = A,\ \mathrm{trace}(A) = m,\ A\mathbf{1}_k = 0\} \tag{7.2}$$

which is a compact Riemannian submanifold of $S(k,\mathbb{R})$ of dimension $mk - m - m^2$. It is easy to show that $A = u'(uu')^{-1}u$.

Let Q be a probability distribution on $A\Sigma_m^k$ and let $\tilde{Q} = Q \circ J^{-1}$ be its image in $J(A\Sigma_m^k)$. Let $\tilde{\mu}$ be the mean of $\tilde{Q}$, that is $\tilde{\mu} = \int_{J(A\Sigma_m^k)} x\tilde{Q}(dx)$. Then $\tilde{\mu}$ is a $k \times k$ positive semi definite matrix satisfying

$$\mathrm{trace}(\tilde{\mu}) = m,\ \mathrm{rank}(\tilde{\mu}) \geq m \text{ and } \tilde{\mu}\mathbf{1}_k = 0.$$

Let $P(\tilde{\mu})$ be the set of projections of $\tilde{\mu}$ into $J(A\Sigma_m^k)$, as described in (3.3). Proposition 7.1 below gives an expression for $P(\tilde{\mu})$ and hence finds the extrinsic mean set of Q. It has been proved in Sughatadasa (2006).

PROPOSITION 7.1. *(a) The projection of $\tilde{\mu}$ into $J(A\Sigma_m^k)$ is given by*

$$P(\tilde{\mu}) = \{\sum_{j=1}^{m} U_j U_j'\} \tag{7.3}$$

where $U = (U_1, \ldots, U_k) \in SO(k)$ is such that $\tilde{\mu} = U\Lambda U'$, $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_k)$, $\lambda_1 \geq \ldots \geq \lambda_k = 0$. (b) $\tilde{\mu}$ is nonfocal and Q has a unique extrinsic mean μ_E iff $\lambda_m > \lambda_{m+1}$. Then $\mu_E = \sigma(F')$ where $F = (U_1, \ldots, U_m)$.

We can use Proposition 7.1 and Proposition 3.1 to get an expression for the extrinsic variation V of Q as follows:

$$\begin{aligned} V &= \|\tilde{\mu} - P(\tilde{\mu})\|^2 + \int_{J(A\Sigma_m^k)} \|\tilde{\mu} - x\|^2 \tilde{Q}(dx) \\ &= 2(m - \sum_{i=1}^{m} \lambda_i). \end{aligned} \tag{7.4}$$

Let $X_1, \ldots, X_n$ be an iid sample from Q and let μ_{nE} be the sample extrinsic mean, which is a measurable selection from the sample extrinsic mean set. It follows from Proposition 2.2 that if Q has a unique extrinsic mean μ_E, that is, if $\tilde{\mu}$ is a nonfocal point of $S(k, \mathbb{R})$, then μ_{nE} is a consistent estimator of μ_E.

7.1. Asymptotic distribution of the sample extrinsic mean. In this section, we assume that $\tilde{\mu}$ is a nonfocal point of $S(k, \mathbb{R})$. Then the map $P(\tilde{\mu}) = \sum_{j=1}^{m} U_j U_j'$ is well defined and smooth in a neighborhood $N(\tilde{\mu})$ of $\tilde{\mu}$ in $S(k, \mathbb{R})$. That follows from Perturbation theory, because if $\lambda_m > \lambda_{m+1}$, then the subspace spanned by $\{U_1, \ldots, U_m\}$ is a smooth map into $G_m(k-1)$ and $P(\tilde{\mu})$ is the projection matrix onto that subspace. Then $\sqrt{n}(J(\mu_{nE}) - J(\mu_E))$ is asymptotically normal in the tangent space of $J(A\Sigma_m^k)$ at $J(\mu_E) \equiv P(\tilde{\mu})$. To get the asymptotic dispersion, we need to find the derivative of P. Define

$$N_m^k = \{A \in S(k, \mathbb{R})\colon A^2 = A, \mathrm{trace}(A) = m\}. \tag{7.5}$$

Then $N_m^k = J(A\Sigma_m^{k+1})$, which is a Riemannian manifold of dimension $km - m^2$. It has been shown in Dimitric (1996) that the tangent and normal spaces to N_m^k are given by

$$T_A N_m^k = \{v \in S(k, \mathbb{R})\colon vA + Av = v\}, \tag{7.6}$$

$$T_A {N_m^k}^{\perp} = \{v \in S(k, \mathbb{R})\colon vA = Av\}. \tag{7.7}$$

Consider the map

$$P : N(\tilde{\mu}) \to N_m^k,\ P(\mu) = \sum_{j=1}^{m} U_j(\mu)U_j(\mu)' \tag{7.8}$$

where $\mu = \sum_{j=1}^{k} \lambda_j(\mu)U_j(\mu)U_j(\mu)'$ is a s.v.d. of μ as in Proposition 7.1.

Proposition 7.2. *The derivative of P is given by*

$$dP : S(k,\mathbb{R}) \to TN_m^k,\ d_{\tilde{\mu}}P(A) = \sum_{i=1}^{m} \sum_{j=m+1}^{k} (\lambda_i - \lambda_j)^{-1} a_{ij} U E_{ij} U' \tag{7.9}$$

where $A = \sum\sum_{1\le i\le j\le k} a_{ij} U E_{ij} U'$ and $\{UE_{ij}U' : 1 \le i \le j \le k\}$ is the orthonormal basis (frame) for $S(k,\mathbb{R})$ mentioned in Section 6.1.

Proof. Let $\gamma(t) = \tilde{\mu} + tv$ be a curve in $N(\tilde{\mu})$ with $\gamma(0) = \tilde{\mu}$ and $\dot{\gamma}(0) = v \in S(k,\mathbb{R})$. Then

$$\gamma(t) = U(\Lambda + tU'vU)U' = U\tilde{\gamma}(t)U' \tag{7.10}$$

where $\tilde{\gamma}(t) = \Lambda + tU'vU$, which is a curve in $S(k,\mathbb{R})$ satisfying $\tilde{\gamma}(0) = \Lambda$ and $\dot{\tilde{\gamma}}(0) = \tilde{v} = U'vU$. From (7.8) and (7.10), we get

$$P[\gamma(t)] = UP[\tilde{\gamma}(t)]U'. \tag{7.11}$$

Differentiate (7.11) at $t = 0$ to get

$$d_{\tilde{\mu}}P(v) = Ud_{\Lambda}P(\tilde{v})U'. \tag{7.12}$$

To find $d_{\Lambda}P(\tilde{v}) = \frac{\mathrm{d}}{\mathrm{d}t}P[\tilde{\gamma}(t)]|_{t=0}$, we may assume w.l.o.g. that $\lambda_1 > \lambda_2 > \ldots > \lambda_k$. Then we can choose a s.v.d. for $\tilde{\gamma}(t)$ as $\tilde{\gamma}(t) = \sum_{j=1}^{k} \lambda_j(t)e_j(t)e_j(t)'$ such that $\{e_j(t), \lambda_j(t)\}_{j=1}^{k}$ are some smooth functions of t satisfying $e_j(0) = e_j$ and $\lambda_j(0) = \lambda_j$, where $\{e_j\}_{j=1}^{k}$ is the canonical basis for $\mathbb{R}^k$. Let $\tilde{v} = E_{ab}$, $1 \le a \le b \le k$. Then we can get expressions for $\dot{e}_j(0)$ from (6.36). Since

$$P[\tilde{\gamma}(t)] = \sum_{j=1}^{m} e_j(t)e_j(t)',$$

therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}P[\tilde{\gamma}(t)]|_{t=0} = \sum_{j=1}^{m} [e_j\dot{e}_j(0)' + \dot{e}_j(0)e_j'] \tag{7.13}$$

From (6.36) and (7.13), we get that

$$d_\Lambda P(E_{ab}) = \begin{cases} (\lambda_a - \lambda_b)^{-1} E_{ab} & \text{if } a \le m < b \le k, \\ 0 & \text{otherwise} \end{cases} \tag{7.14}$$

Then from (7.12), we get that

$$d_{\tilde{\mu}} P(UE_{ab}U') = \begin{cases} (\lambda_a - \lambda_b)^{-1} UE_{ab}U' & \text{if } a \le m < b \le k, \\ 0 & \text{otherwise} \end{cases} \tag{7.15}$$

Hence if $A = \sum\sum_{1 \le i \le j \le k} a_{ij} UE_{ij}U'$, from (7.15), we get

$$d_{\tilde{\mu}} P(A) = \sum_{i=1}^{m} \sum_{j=m+1}^{k} (\lambda_i - \lambda_j)^{-1} a_{ij} UE_{ij}U'. \tag{7.16}$$

This completes the proof. □

Corollary 7.1. *Consider the projection map of* (7.8) *restricted to*

$$S_0(k, \mathbb{R}) := \{A \in S(k, \mathbb{R}) : A\mathbf{1}_k = 0\}.$$

It has the derivative

$$dP : S_0(k, \mathbb{R}) \to TJ(A_m^k), \ d_{\tilde{\mu}} P(A) = \sum_{i=1}^{m} \sum_{j=m+1}^{k-1} (\lambda_i - \lambda_j)^{-1} a_{ij} UE_{ij}U'. \tag{7.17}$$

Proof. Follows from Proposition 7.2 and the fact that

$$T_{P(\tilde{\mu})} J(A_m^k) = \{v \in T_{P(\tilde{\mu})} N_m^k \colon v\mathbf{1}_k = 0\}.$$

□

From Corollary 7.1, it follows that $\{UE_{ij}U' : 1 \le i \le m < j < k\}$ forms an orthonormal basis for $T_{P(\tilde{\mu})} J(A_m^k)$ and if $A \in S(k, \mathbb{R})$ has coordinates $\{a_{ij} : 1 \le i \le j \le k\}$ w.r.t the orthonormal basis $\{UE_{ij}U' : 1 \le i \le j \le k\}$ of $S(k, \mathbb{R})$, then $d_{\tilde{\mu}} P(A)$ has coordinates $\{(\lambda_i - \lambda_j)^{-1} a_{ij} : 1 \le i \le m < j < k\}$ in $T_{P(\tilde{\mu})} J(A_m^k)$. Also it is easy to show that the linear projection $L(A)$ of A into $T_{P(\tilde{\mu})} J(A_m^k)$ has coordinates $\{a_{ij} : 1 \le i \le m < j < k\}$.

Now we can get asymptotic coordinates for $\sqrt{n}[J(\mu_{nE}) - J(\mu_E)]$ as in Section 3.1 and perform two sample tests as in Section 3. An expression for the covariance matrix as in (3.7) may also be found in Patrangenaru and Sughatadasa (2005).

8 Applications to Shape Data

In this section, we mention two applications where we use the methods developed in the earlier sections to identify differences between extrinsic means and variations of two populations of landmark based shapes. The numerical computations are done using Matlab software.

8.1. Application to Glaucoma detection. To detect any shape change due to Glaucoma, 3D images of the Optic Nerve Head (ONH) of both eyes of 12 mature rhesus monkeys were collected. One of the eyes was treated to increase the Intra Ocular Pressure (IOP) which is often the case of glaucoma onset, while the other was left untreated. 5 landmarks were recorded on each eye. For details on landmark registration, see Derado et al. (2004). The landmark coordinates can be found in BP (2005). In this section, we consider the reflection shape of the k-ads in $R\Sigma_3^k$, $k = 5$. We want to test if there is any significant difference between the shapes of the treated and untreated eyes by comparing the extrinsic means and variations.

Figure 1(a) shows the preshapes of the k-ads from the untreated eyes along with a preshape of the sample extrinsic mean. The five landmarks are labelled from 1 to 5. The sample preshapes have been rotated and (or) reflected so as to minimize their Euclidean distance from the mean's preshape. Figure 1(b) shows the corresponding landmarks for the k-ads from the treated eyes along with those from the sample extrinsic mean. Again seperate rotation/reflection have been applied to each sample preshape so as to bring it closest to the preshape of the mean. Figure 2 shows the preshapes of the sample extrinsic means for the two eyes along with a preshape of the pooled sample extrinsic mean. Again the two means' preshapes have been rotated/reflected appropriately to bring them closest to the pooled sample mean's preshape. The sample extrinsic means have coordinates

$$L[P(\hat{\mu}_1) - P(\hat{\mu})] = (0.003, -0.011, -0.04, 0.021, 0.001, -0.001, 0.007, -0.004),$$
$$L[P(\hat{\mu}_2) - P(\hat{\mu})] = (-0.003, 0.011, 0.04, -0.021, -0.001, 0.001, -0.007, 0.005)$$

in the tangent space at $P(\hat{\mu})$. Here $P(\hat{\mu}_1)$ and $P(\hat{\mu}_2)$ are the embeddings of the sample extrinsic mean shapes of the untreated and treated eyes respectively, $P(\hat{\mu})$ is the embedded extrinsic mean shape of the pooled sample and L denotes the linear projection into $T_{P(\hat{\mu})}J(R\Sigma_3^5)$. The sample extrinsic variations for the untreated and treated eyes are 0.041 and 0.038 respectively.

The value of the matched pair test statistic T_{1p} in Section 3.2.2 is 36.29 and the asymptotic p-value for testing if the shape distributions for the two eyes are the same is

$$P(\mathcal{X}_8^2 > 36.29) = 1.55 \times 10^{-5}.$$

The value of the test statistic T_{2p} for testing whether the extrinsic means are the same is 36.56 and the p-value of the chi-squared test turns out to be 1.38×10^{-5}. Hence we conclude at asymptotic level 0.0001 or higher that the mean shapes of the two eyes are significantly different. Because of lack of sufficient data and high dimension, the bootstrap estimates of the covariance matrix $\hat{\Sigma}$ in (3.40) turn out to be singular or close to singular in many simulations. To avoid that, we construct a pivotal bootsrap confidence region for the first few principal scores of $L_{\tilde{\mu}}[P(\mu_1) - P(\mu_2)]$ and see if it includes $\mathbf{0}$. Here $P(\mu_i)$ is the embedding of the extrinsic mean of Q_i, $i = 1, 2$ (see Section 3.2.2) and $\tilde{\mu} = (\mu_1 + \mu_2)/2$. The first two principal components of $\hat{\Sigma}$ explain more than 80% of its variation. A confidence region for the first two principal scores is given by the inequality

$$nT_n'\hat{\Sigma}_{11}^{-1}T_n \le c^*(1-\alpha) \text{ where} \tag{8.1}$$

$$T_n = L[P(\hat{\mu}_1) - P(\hat{\mu}_2) - P(\mu_1) + P(\mu_2)]. \tag{8.2}$$

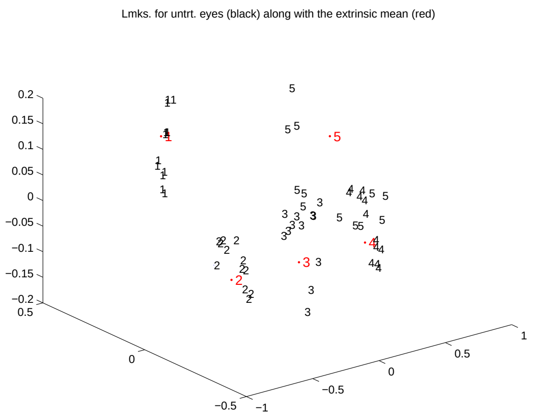
Here $n = 12$ is the sample size and $c^*(1-\alpha)$ is the $(1-\alpha)$-quantile of the bootstrap distribution of $nv^*\Sigma_{11}^{*-1}v^*$, v^* being defined in (3.29). If $\hat{\Sigma} = \sum_{j=1}^{8} \lambda_j U_j U_j'$ be a s.v.d. for $\hat{\Sigma}$, then $\hat{\Sigma}_{11}^{-1} \doteq \sum_{j=1}^{2} \lambda_j^{-1} U_j U_j'$ and Σ_{11}^{*-1} is its bootstrap estimate. The bootstrap p-value using 10^4 resamples turns out to be 0.0098. Hence we again reject $H_0 : P(\mu_1) = P(\mu_2)$. The corresponding p-value using $\mathcal{X}_2^2$ approximation for the distribution of $nT_n'\hat{\Sigma}_{11}^{-1}T_n$ in (8.1) turns out to be 0.002.

Two sample tests to compare the mean shapes have also been carried out in BP (2005) and Bandulasiri et al. (2007) by different methods but those tests yield much bigger p-values.

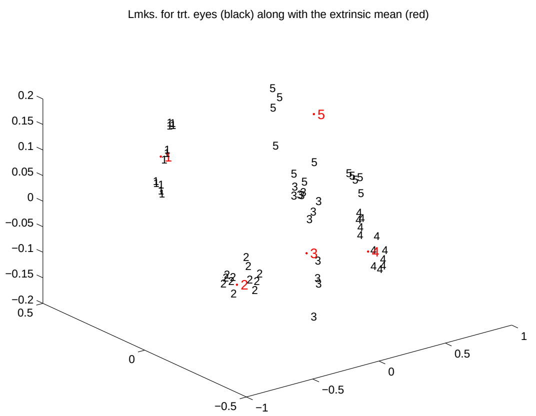
Next we test if the two eye shapes have the same extrinsic variation. The value of the test statistic T_{3p} in (3.44) equals -0.5572 and the asymptotic p-value equals

$$P(|Z| > 0.5572) = 0.577, \ Z \sim N(0,1).$$

The bootstrap p-value from 10^4 resamples equals 0.59. Hence we accept H_0 and conclude that the extrinsic variations are equal at levels 0.5 or lower.



(a)



(b)

Figure 1: (a) and (b) show 5 landmarks from untreated and treated eyes of 12 monkeys, respectively, along with the sample extrinsic mean shapes.. correspond to the mean shapes' landmarks (labelled from 1 to 5).

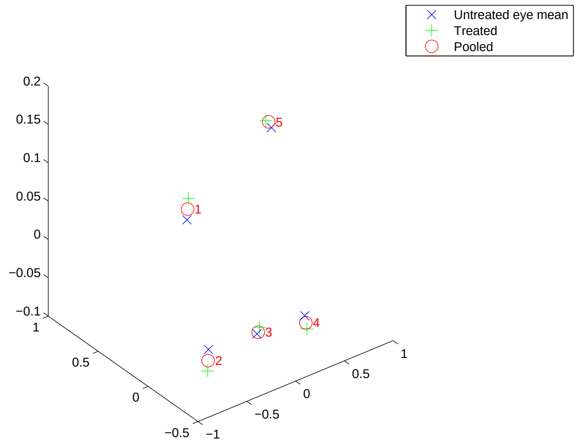


Figure 2: Landmarks from the sample extrinsic mean shapes for the 2 eyes along with those from the pooled sample extrinsic mean (labelled from 1 to 5).

8.2. Application to handwritten digit recognition. A random sample of 30 handwritten digit '3' were collected so as to devise a scheme to automatically classify handwritten characters. 13 landmarks were recorded on each image by Anderson (1997). The landmark data can be found in Dryden and Mardia (1998).

We analyze the affine shape of the sample points and estimate the mean shape and variation in shape. This can be used as a prior model for digit recognition from images of handwritten codes. Our observations lie on the affine shape space $A\Sigma_2^k$, $k = 13$. A representative of the sample extrinsic mean shape μ_{nE} has coordinates

$$u = (-0.53, -0.32, -0.26, -0.41, 0.14, -0.43, 0.38, -0.29, 0.29, -0.11, 0.06, 0,$$
$$-0.22, 0.06, 0.02, 0.08, 0.19, 0.13, 0.30, 0.21, 0.18, 0.31, -0.13, 0.38, -0.42, 0.38)$$

so that $\mu_{nE} = \sigma(u)$. The coordinates are in pairs, x coordinate followed by y. Figure 3 shows the plot of u. The sample extrinsic variation turns out to be 0.27 which is fairly large. There seems to be a lot of variability in the data. Following are the extrinsic distances squared of the sample points from the mean shape:

$$(\rho^2(X_j, \mu_{nE}),\ j = 1, \ldots, n) = (1.64, 0.28, 1.00, 0.14, 0.13, 0.07, 0.20, 0.09,$$
$$0.17, 0.15, 0.26, 0.17, 0.14, 0.20, 0.42, 0.31, 0.14, 0.12, 0.51, 0.10, 0.06, 0.15$$
$$0.05, 0.31, 0.08, 0.08, 0.11, 0.18, 0.64, 0.12).$$

Here $n = 30$ is the sample size. From these distances, it is clear that observations 1 and 3 are outliers. We remove them and recompute the sample extrinsic mean and variation. The sample variation now turns out to be 0.19. An asymptotic 95% confidence region for the extrinsic mean μ_E as in (3.9) is given by

$$\{\mu_E : n(L[P(\tilde{\mu}) - P(\bar{Y})])'\hat{\Sigma}^{-1}L[P(\tilde{\mu}) - P(\bar{Y})] \leq \mathcal{X}^2_{20}(0.95) = 31.4104\}.$$

The dimension 20 of $A\Sigma_2^{13}$ is quite high compared to the sample size of 28. It is difficult to construct a pivotal bootstrap confidence region as in (3.10) because the bootstrap covariance estimates Σ^* tend to be singular or close to singular in most simulations. Instead, we construct the following nonpivotal bootstrap confidence region:

$$\{\mu_E : n(L[P(\tilde{\mu}) - P(\bar{Y})])'\hat{\Sigma}^{-1}L[P(\tilde{\mu}) - P(\bar{Y})] \leq c^{**}(1-\alpha)\}.$$

where $c^{**}(1-\alpha)$ is the $(1-\alpha)$-quantile of the bootstrap distribution of

$$n(L[P(\bar{Y}) - P(\bar{Y}^*)])'\hat{\Sigma}^{-1}L[P(\bar{Y}) - P(\bar{Y}^*)].$$

Then the 95^{th} bootstrap percentile $c^{**}(0.95)$ turns out be 1.077 using 10^5 resamples. Hence bootstrap methods yield much smaller confidence region for the true mean shape compared to that obtained from chi-squared approximation.

A 95% confidence interval for the extrinsic variation V by normal approximation is given by $V \in [0.140, 0.243]$ while a pivotal bootstrap confidence interval using 10^5 bootstrap resamples turns out to be $[0.119, 0.264]$.

Acknowledgment. This article constitutes a part of the author's PhD dissertation. The author would like to thank his advisor, Professor R. Bhattacharya for his kind and helpful suggestions while writing the article. The author is also grateful to the reviewer for going over the article, and sending a highly positive feedback and valuable comments towards improving its contents.

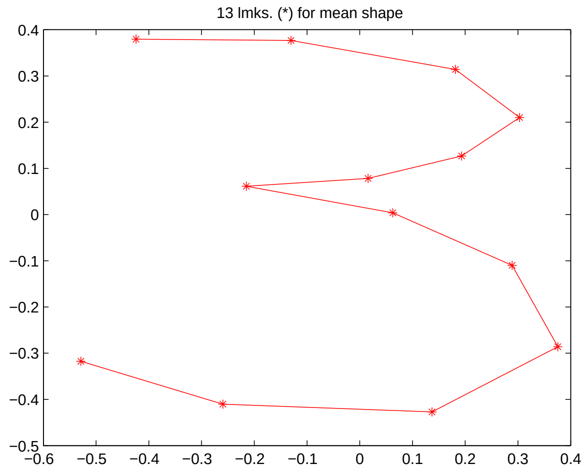


Figure 3: Sample extrinsic mean shape for handwritten digit 3 sample.

References

Anderson, C. R. (1997). *Object recognition using statistical shape analysis.* PhD thesis, University of Leeds.

Bandulasiri, A., Bhattacharya, R. N. and Patrangenaru, V. (2007). Nonparametric Inference on Shape Manifolds with Applications in Medical Imaging. *To appear.*

Bandulasiri A. and Patrangenaru, V. (2005). Algorithms for Nonparametric Inference on Shape Manifolds, *Proc. of JSM 2005, Minneapolis, MN* 1617-1622.

Berthilsson, R. and Heyden A. (1999). Recognition of Planar Objects using the Density of Affine Shape. *Computer Vision and Image Understanding*, **76**, 135-145.

Berthilsson, Rikard and Astrom, K. (1999). Extension of affine shape. *J. Math. Imaging Vision*, **11**, 119-136.

Bhattacharya, A. and Bhattacharya, R. (2008a). Nonparametric Statistics on Manifolds with Applications to Shape Spaces. *Pushing the Limits of Contemporary Statistics: Contributions in honor of J.K. Ghosh. IMS Collections*, **3** 282-301.

Bhattacharya, A. and Bhattacharya, R. (2008b). Statistics on Riemannian Manifolds: Asymptotic Distribution and Curvature. *Proc. Amer. Math. Soc*, **136**, 2957-2967.

Bhattacharya, R. and Bhattacharya, A. (2008c). Statistics on Manifolds with Applications to Shape Spaces. *Perspectives in Mathematical Sciences.* Indian Statistical Institute, Bangalore. *In Press.*

Bhattacharya, R. N. and Patrangenaru, V. (2002). Nonparametric estimation of location and dispersion on Riemannian manifolds. *J. Statist. Plann. Inference*, **108** 23-35.

BHATTACHARYA, R. N. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds-I. *Ann. Statist.*, **31** 1-29.

BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds-II. *Ann. Statist.*, **33** 1225-1259.

BOOKSTEIN, F. (1978). *The Measurement of Biological Shape and Shape Change.* Lecture Notes in Biomathematics, **24**. Springer, Berlin.

BOOKSTEIN, F. L. (1986). Size and shape spaces of landmark data (with discussion). *Statistical Science*, **1**, 181-242.

BOOKSTEIN, F.L. (1997). *Morphometric Tools for Landmark data: Geometry and Biology.* Cambridge Univ. Press, Cambridge, U.K.

BOOTHBY, W. (1986). *An Introduction to Differentiable Manifolds and Riemannian Geometry.* 2d ed. Academic Press, Orlando.

DERADO, G., MARDIA, K.V., PATRANGENARU, V. and THOMPSON, H.W. (2004). A Shape Based Glaucoma Index for Tomographic Images. *J. Appl. Statist.*, **31** 1241-1248.

DIMITRIĆ, IVKO (1996). A note on equivariant embeddings of Grassmannians. *Publ. Inst. Math.* (Beograd) (N.S.), **59**, 131-137.

DO CARMO, M. P. (1992). *Riemannian Geometry.* Birkhauser, Boston. English translation by F. Flaherty.

DRYDEN, I. L., KUME, A., LE, H. and WOOD, A. T.A. (2008). A multidimensional scaling approach to shape analysis. *Biometrika*, **95**, 779-798.

DRYDEN, I. L. and MARDIA, K. V. (1998). *Statistical Shape Analysis.* Wiley N.Y.

DUNFORD, N. and SCHWARTZ, J. (1958). *Linear Operators - I.* Wiley N.Y.

FRÉCHET, M. (1948). Lés élements aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, **10** 215-310.

HENDRICKS, H. and LANDSMAN, Z. (1988). Mean location and sample mean location on maniforls. Asymptotic, tests, confidence regions. *J. Multivariate Anal.*, **67** 227-243.

HENDRICKS, H. and LANDSMAN, Z. (2007). Asymptotic data analysis on manifolds. *Ann. Statist.*, **35**, 109-131.

KENDALL, D.G. (1977). The diffusion of shape. *Adv. Appl. Probab.*, **9**, 428-430.

KENDALL, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.*, **16**, 81-121.

KENDALL, DAVID G. (1989) A survey of the statistical theory of shape. *Statist. Sci.*, **4**, 87-120.

KENDALL, D. G., BARDEN, D., CARNE, T. K. and LE, H. (1999). *Shape & Shape Theory.* Wiley N.Y.

MARDIA, K.V. and PATRANGENARU, V. (2005). Directions and projective shapes. *Ann. Statist.*, **33**, 1666-1699.

MARDIA, K.V., PATRANGENARU, V. and SUGATHADASA, S. (2005). Protein Gels Matching. *Quantitative Biology, Shape Analysis and Wavelets.* 163-165. Leeds, Leeds Uni. Press.

PATRANGENARU, V. (1998). *Asymptotic Statistics on Manifolds and Their Applications*, PhD. dissertation, Indiana University, Bloomington.

PATRANGENARU, V. AND SUGHATADASA, S. (2005). A Covariance Formula for Shape Statistics on Grassmannians. *Proceedings of ICIA05 Conference.* 441-445. Colombo, Sri Lanka.

Sepiashvili, D., Moura, J.M.F. and Ha, V.H.S. (2003). Affine-Permutation Symmetry: Invariance and Shape Space (2003). *Proceedings of the 2003 Workshop on Statistical Signal Processing*. 293-296.

Small, C.G. (1996). *The Statistical Theory of Shape*. Springer, New York.

Sparr, G. (1992). Depth-computations from polihedral images. *Proc. 2nd European Conf. on Computer Vision*(ECCV-2) (G. Sandini, editor). 378-386.

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

Abhishek Bhattacharya
Department of Statistical Science
Duke University, Box 90251
Durham, NC 27708
E-mail: abhishek.bhattacharya@duke.edu

Paper received October 2008; revised January 2009.