STATISTICS ON RIEMANNIAN MANIFOLDS WITH APPLICATIONS TO THE PLANER SHAPE SPACE

Submitted by: ABHISHEK BHATTACHARYA
Project Advisor: Dr. RABI BHATTACHARYA

November 15, 2006

Abstract

This article presents certain recent methodologies and some new results for the statistical analysis of distributions of shapes on manifolds. An important example considered in some detail here is the 2-D shape space of k-ads, comprising all configurations of k planar landmarks ($k > 2$) -modulo translation, scaling and rotation. If one leaves out configurations of identical k points, the planar shape space can be identified with the complex projective space $\mathbb{C}P^{k-2}$.

The statistical analysis of shape distributions based on random samples is important in many areas such as morphometrics (discrimination and classification of biological shapes), medical diagnostics (detection of change or deformation of shapes in some organs due to some disease, for example) and machine vision (e.g., digital recording and analysis based on planar views of 3-D objects, when the position from which the object was viewed or pictured is unknown).

1 Frechet Mean and Variation on Metric Spaces.

Let (M, d) be a metric space and Q a probability measure on M. Define the **Frechet function** of Q as

$$F(p) = \int_M d^2(p, x)Q(dx), \quad p \in M. \tag{1.1}$$

1.1 Frechet Mean

Definition 1.1 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 **Frechet Mean set** of Q. If this set is a singleton, that is the **Frechet Mean** of Q.
If $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) with commom distribution Q, and $Q_n \doteq \frac{1}{n}\sum_{j=1}^{n} \delta_{X_j}$ is the corresponding empirical distribution, then the Frechet mean set of Q_n is called the **sample Frechet mean set**. If this set is a singleton, it is called the **sample Frechet mean**.

Proposition 1.1 Suppose every closed and bounded subset of M is compact. If the Frechet function $F(p)$ of Q is finite for some p, then the Frechet mean set of Q is nonempty and compact.
Proof By the triangle inequality,

$$\begin{aligned} d^2(q,x) &\leq (d(p,q) + d(p,x))^2 \\ &= d^2(p,x) + d^2(p,q) + 2d(p,q)d(p,x) \end{aligned}$$

Hence

$$\begin{aligned} &F(q) \leq F(p) + d^2(p,q) + 2d(p,q)F^{1/2}(p), \\ &|F(q) - F(p)| \leq d^2(p,q) + 2d(p,q)\max\{F^{1/2}(p), F^{1/2}(q)\} \end{aligned} \tag{1.2}$$

Hence if $F(p)$ is finite for some p, it is finite for all p, and $p \mapsto F(p)$ is continuous on M. To show that a minimizer exists when F is finite, let c denote the infimum of F, and let $\{p_n : n \geq 1\}$ be a sequence such that $F(p_n) \longrightarrow c$. Now

$$\begin{aligned} &d^2(p_n, p_1) \leq 2d^2(p_n, x) + 2d^2(x, p_1), \ x \in M. \\ \Rightarrow &d^2(p_n, p_1) \leq 2F(p_n) + 2F(p_1) \end{aligned} \tag{1.3}$$

proving that $B = \{p_n : n \geq 1\}$ is a bounded sequence, so that its closure $\bar{B}$ is compact (Note that $\text{diam}(\bar{B}) = \text{diam}(B) < \infty$). Then $\{p_n : n \geq 1\}$ has a convergent subsequence $p_{n_k} \longrightarrow p^*$ as $k \to \infty$. By continuity of F, $F(p^*) = c$, so that p^* is a minimizer of F. Also if m is an arbitrary minimizer, $d^2(m, p^*) \leq 2F(m) + 2F(p^*) < 4c$. Hence the set of all minimizers is a bounded set, say, D. Since every point in $\bar{D}$ is a limit of a sequence in D, and F is continuous, every point in $\bar{D}$ is a minimizer of F. Hence $D = \bar{D}$ is compact. $\square$

Theorem 1.2 **(Consistency of the Sample Frechet Mean)**. Assume (1) that every closed bounded subset of M is compact, and (2) F is finite on M. Then (a) given any $\epsilon > 0$, there exists a P-null set N and $n(\omega) < \infty \forall \omega \in N^c$, such that the Frechet sample mean set of the empirical distribution $Q_n = Q_n(\omega)$ is contained in the $\epsilon-$neighborhood of the Frechet mean set of Q $\forall n \geq n(\omega)$, and (b) if the Frechet mean of Q exists (as a unique minimizer of F), then every measurable selection from the Frechet sample mean set is a strongly consistent estimator of the Frechet mean of Q.
Proof See Bhattacharya & Patrangenaru Theorem 2.3 [1]. $\square$

Note that a sequence of estimator θ_n defined on a probability space $(\Omega, P, \mathbb{B})$ is said to be a strongly consistent estimator of a parameter θ, if $\theta_n(\omega) \longrightarrow \theta$ as $n \rightarrow \infty$ for every ω outside of a P-null set.

Remark 1.1 It is known that a connected Riemannian manifold, M which is complete (in its geodesic distance) satisfies the topological hypothesis of Proposition 1.1: every closed bounded subset of M is compact. We will see sufficient conditions for the existence of the Frechet mean of Q (as a unique minimizer of the Frechet function F of Q) in the subsequent sections.

Now we deduce the asymptotic distribution of the sample mean after proper scaling and translation in case there is a population mean. This result can be used to construct asymptotic confidence set for the population mean based on the sample analogue and for non-parametric testing. We shall see these applications in the Planer Shape Space.

Theorem 1.3 **(Asymptotic distribution of Frechet Sample mean)** Suppose the following assumptions hold:

A1 Q has support in a single coordinate patch, (U, ϕ), $\phi : U \longrightarrow \mathbb{R}^d$ smooth. Let $Y_j = \phi(X_j)$; $j = 1, \ldots, n$. $Q^\phi = Q \circ \phi^{-1}$.

A2 Frechet Mean μ_F of Q is unique.

A3 $\forall x,\ y \mapsto h(x, y) = (d^\phi)^2(x, y) = d^2(\phi^{-1}x, \phi^{-1}y)$ is twice continuously differentiable in a neighborhood of $\phi(\mu_F) = \mu$.

A4 $E(D_r h(Y,\mu))^2 < \infty \ \forall r$.

A5 $E\{\sup_{|u-v|\leq\epsilon} |D_s D_r h(Y_1, v) - D_s D_r h(Y_1, u)|\} \to 0$ as $\epsilon \to 0 \ \forall \ r, s$.

A6 $\Lambda = ((\ E D_s D_r h(Y_1, \mu)\ ))$ is nonsingular.

A7 $\Sigma = \mathrm{Cov}\ Dh(Y_1, \mu)$ is nonsingular.
Let $\mu_{F,n}$ be a measurable selection from the Frechet sample mean set, $\mu_n = \phi(\mu_{F,n})$. Then under the assumptions A1-A7,

$$\sqrt{n}(\mu_n - \mu) \xrightarrow{\mathcal{L}} N(0, \Lambda^{-1}\Sigma(\Lambda')^{-1})$$

<u>Proof</u> Let $F(y) = \int {d^{\phi}}^2(x,y) Q^{\phi}(dx)$. Its minimizer is μ.
Similarly define $F_n(y) = \int {d^{\phi}}^2(x,y) Q_n^{\phi}(dx) = \frac{1}{n}\sum_{j=1}^n {d^{\phi}}^2(Y_j, y)$ ($Q_n^{\phi} = \frac{1}{n}\sum_{j=1}^n \delta_{Y_j}$).
Its minimizer is μ_n. So,

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}}\sum_{j=1}^n D_r h(Y_j, \mu_n) \\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^n D_r h(Y_j, \mu) \\ &+ \sum_{s=1}^d \sqrt{n}(\mu_n - \mu)^{(s)} \frac{1}{n}\sum_{j=1}^n D_s D_r h(Y_j, \mu) + \sum_{s=1}^d \sqrt{n}(\mu_n - \mu)^{(s)} (\epsilon_n)_{rs}, \ 1 \leq r \leq d \end{aligned} \tag{1.4}$$

where $(\epsilon_n)_{rs} = \dfrac{1}{n}\displaystyle\sum_{j=1}^n [D_s D_r h(Y_j, \theta_n) - D_s D_r h(Y_j, \mu)]$

for some θ_n lying on the line segment joining μ and μ_n.

$$\Rightarrow \left[\left(\left(\frac{1}{n}\sum_{j=1}^n D_s D_r h(Y_j, \mu) + \epsilon_n\right)\right)\right] \sqrt{n}(\mu_n - \mu) = -\frac{1}{\sqrt{n}}\sum_{j=1}^n Dh(Y_j, \mu)$$

$$\Rightarrow \sqrt{n}(\mu_n - \mu) = -\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^n Dh(Y_j, \mu)\right) + o_P(1) \tag{1.5}$$

$$\Rightarrow \sqrt{n}(\mu_n - \mu) \xrightarrow{\mathcal{L}} -\Lambda^{-1} N(0, \Sigma) = N(0, \Lambda^{-1}\Sigma(\Lambda')^{-1}) \tag{1.6}$$

□

1.2 Frechet Variation

Definition 1.2 The **Frechet Variation**, V of Q is the minimum value attained by the Frechet function F on M. Similarly the minimum value attained by the sample Frechet function,

$$F_n(p) = \frac{1}{n}\sum_{i=1}^{n} d^2(X_i, p)$$

is called the **sample Frechet Variation**.

From Proposition 1.1., it follows that if the Frechet function is finite for some p, then the variation is finite and is attained by all p in the Frechet mean set. Similarly the sample variation is the value of F_n on the sample mean set.

Theorem 1.4 **(Consistency of the Sample Variation)**Suppose every closed and bounded subset of M is compact, and F is finite on M. Then the sample Frechet variation is a strongly consistent estimator of the population variation.
Proof Let $\hat{\theta}_n$ be a measurable selection from the sample Frechet mean set. Then the sample variation $V_n = F_n(\hat{\theta}_n)$.
Fix $\epsilon > 0$.
By Theorem 1.2, there exists a P-null set A and for all $\omega \in A^c$, there exists $N(\omega) < \infty$, such that for all $n \geq N(\omega)$, $\hat{\theta}_n(\omega)$ is contained in the $\epsilon-$neighborhood of the Frechet mean set of Q.
From Proposition 1.1, the Frechet mean set is compact. So we may assume that $\forall\ n \geq N(\omega)$, $\hat{\theta}_n(\omega) \in K$, where K is a compact set containing the Frechet mean set of Q. Choose θ_n in the Frechet mean set. Then

$$\begin{aligned}
|V_n - V| &= \left|\frac{1}{n}\sum_{j=1}^{n} d^2(X_j, \hat{\theta}_n) - Ed^2(X_1, \theta_n)\right| \\
&\leq \left|\frac{1}{n}\sum_{j=1}^{n}(d^2(X_j, \hat{\theta}_n) - d^2(X_j, \theta_n))\right| + \left|\frac{1}{n}\sum_{j=1}^{n} d^2(X_j, \theta_n) - Ed^2(X_1, \theta_n)\right| \\
&\leq \frac{1}{n}\sum_{j=1}^{n} |d^2(X_j, \hat{\theta}_n) - d^2(X_j, \theta_n)| + \sup_{\theta\in K}\left|\frac{1}{n}\sum_{j=1}^{n} d^2(X_j, \theta) - Ed^2(X_1, \theta)\right|
\end{aligned} \tag{1.7}$$

By Theorem 2.3[1], the second term on the RHS of (1.7) goes to 0. Lets assume that it is less than ϵ, for all $n \geq N(\omega)$.
The first term on RHS of (1.7) is

$$
\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n} |d^2(X_j, \hat{\theta}_n) - d^2(X_j, \theta_n)| &\leq \frac{1}{n}\sum_{j=1}^{n}[d(X_j, \hat{\theta}_n) + d(X_j, \theta_n)]d(\hat{\theta}_n, \theta_n) \\
&\leq \frac{2}{n}\sum_{j=1}^{n} \sup_{\theta \in K} d(X_j, \theta) d(\hat{\theta}_n, \theta_n)
\end{aligned}
$$

$$
\begin{aligned}
\sup_{\theta \in K} d(X_1, \theta) &\leq d(X_1, \theta_0) + \sup_{\theta \in K} d(\theta_0, \theta) \text{ for any } \theta_0 \text{ in } K. \\
\text{So } E \sup_{\theta \in K} d(X_1, \theta) &\leq E\, d(X_1, \theta_0) + \sup_{\theta \in K} d(\theta_0, \theta) < \infty
\end{aligned}
$$

So $\frac{2}{n}\sum_{j=1}^{n} \sup_{\theta \in K} d(X_j, \theta)$ is bounded almost surely, for simplicity lets assume that it is less than 1 for all ω in A^c and $n \geq N(\omega)$.
Choose the sequence θ_n such that $d(\hat{\theta}_n(\omega), \theta_n) < \epsilon$ for $\omega \in A^c$ and $n > N(\omega)$. This is possible by Theorem 1.2. Then the first term on RHS of (1.7) is less than ϵ.
So $|V_n(\omega) - V| < 2\epsilon$ for all $\omega \in A^c$ and $n > N(\omega)$. Since $P(A^c) = 1$, and ϵ is arbitrary, this proves that $V_n \xrightarrow{a.s.} V$. $\square$

Remark 1.2 In the above theorem, we did not need the Frechet mean to exist. The sample variation is a consistent estimator of the true variation even when the Frechet Function of Q does not have a unique minimizer.

Next we deduce the asymptotic distribution of the sample variation when there is a unique population mean. This result fails if we do not have a unique mean.
Theorem 1.5 Assume the same set-up as in Theorem 1.3. Then under assumptions A1-A7,

$$\sqrt{n}(F_n(\mu_n) - F(\mu)) \xrightarrow{\mathcal{L}} N(0, V d^2(X_1, \mu_F))$$

Proof

$$\sqrt{n}(F_n(\mu_n) - F(\mu)) = \sqrt{n}(F_n(\mu_n) - F_n(\mu)) + \sqrt{n}(F_n(\mu) - F(\mu)) \tag{1.8}$$

$$\begin{aligned}\sqrt{n}(F_n(\mu_n) - F_n(\mu)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{r=1}^{d}(\mu_n - \mu)_r D_r h(Y_i, \mu) \\ &\quad + \frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\sum_{r=1}^{d}\sum_{s=1}^{d}(\mu_n - \mu)_r(\mu_n - \mu)_s D_s D_r h(Y_i, \mu_n^*)\end{aligned} \tag{1.9}$$

for some μ_n^* in the line segment joining μ and μ_n.

$$\frac{1}{n}\sum_{i=1}^{n} D_s D_r h(Y_i, \mu_n^*) = \frac{1}{n}\sum_{i=1}^{n} D_s D_r h(Y_i, \mu) + \frac{1}{n}\sum_{i=1}^{n}(D_s D_r h(Y_i, \mu_n^*) - D_s D_r h(Y_i, \mu)) \tag{1.10}$$

Under Assumption (A6), the first term on RHS of (1.10) converges in probability to Λ_{sr} while the second term is $o_P(1)$ (converges to zero in probability) by (A5). So (1.10) is bounded in probability. The second term on the RHS of (1.9) is

$$\frac{1}{2}\sum_{r=1}^{d}\sum_{s=1}^{d}\sqrt{n}(\mu_n - \mu)_r(\mu_n - \mu)_s \frac{1}{n}\sum_{i=1}^{n} D_s D_r h(Y_i, \mu_n^*) \tag{1.11}$$

By Theorem 1.3, $\sqrt{n}(\mu_n - \mu)$ is asymptotically normal, so (1.11) is $o_P(1)$. The first term on the RHS of (1.9) is

$$\langle\sqrt{n}(\mu_n - \mu), \frac{1}{n}\sum_{i=1}^{n} Dh(Y_i, \mu)\rangle \tag{1.10}$$

$\frac{1}{n}\sum_{i=1}^{n} Dh(Y_i, \mu) \to EDh(Y_1, \mu) = 0$ and since $\sqrt{n}(\mu_n - \mu)$ is asymptotically normal, so (1.10) is $o_P(1)$.

This proves that (1.9) is $o_P(1)$. Hence (1.8) becomes

$$\begin{aligned}\sqrt{n}(F_n(\mu_n) - F(\mu)) &= \sqrt{n}(F_n(\mu) - F(\mu)) + o_P(1) \\ F_n(\mu) - F(\mu) &= \frac{1}{n}\sum_{i=1}^{n}\left(d^2(X_i, \mu_F) - Ed^2(X_1, \mu_F)\right) \\ \Rightarrow \sqrt{n}(F_n(\mu_n) - F(\mu)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(d^2(X_i, \mu_F) - Ed^2(X_1, \mu_F)\right)\end{aligned} \tag{1.11}$$

By CLT on the iid sequence $\{d^2(X_j, \mu_F)\}$, (1.11) converges in law to $N(0, Vd^2(X_1, \mu_F))$.
$\square$

From now on let $(\mathbf{M}, \mathbf{g})$ be a d-dimensional connected complete Riemannian manifold, g being the riemannian metric on M. We shall come across different notions of means and variations depending on the distance chosen on M. First we start with the 'Extrinsic distance'.

2 Extrinsic Mean and Variation

Let $\phi : M \to \mathbb{R}^k$ be an isometric map of M onto $\tilde{M} = \phi(M) \subset \mathbb{R}^k$. Define the metric on M as: $d(x, y) = \|\phi(x) - \phi(y)\|$, where $\|.\|$ denotes Euclidean norm ($\|u\|^2 = \sum_{i=1}^{k} {u_i}^2 \ u = (u_1, u_2, .., u_k)$).

Assume $\tilde{M}$ is a closed subset of $\mathbb{R}^k$. Then for every $u \in \mathbb{R}^k$ there exists a compact set of points in $\tilde{M}$ whose distance from u is the smallest among all points in $\tilde{M}$. We will denote this set by $P_{\tilde{M}}u = \{x \in \tilde{M} : \|x - u\| \leq \|y - u\| \ \forall y \in \tilde{M}\}$. If this set is a singleton, u is said to be a **non-focal point** of $\mathbb{R}^k$ (wrt $\tilde{M}$); otherwise it is said to be a **focal point** of $\mathbb{R}^k$.

Definition 2.1 Let (M, d), ϕ be as above. Let Q be a probability measure on M s.t. the Frechet function $F(x) = \int d^2(x, y)Q(dy) < \infty$ $(\forall x)$. The Frechet mean (set) of Q is called the **Extrinsic Mean**(set) of Q, and the Frechet variation of Q is called its **Extrinsic Variation**.
If X_i $(i = 1, \ldots, n)$ are iid observations from Q, and $Q_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ is the empirical distribution, then the Frechet mean(set) of Q_n is called the **Extrinsic Sample mean**(set) and the Frechet variation of Q_n is called the **Extrinsic Sample Variation**.

Let $\tilde{Q}$ and $\tilde{Q}_n$ be the images of Q and Q_n respectively on $\mathbb{R}^k$: $\tilde{Q} = Q \circ \phi^{-1}$, $\tilde{Q}_n = Q_n \circ \phi^{-1}$. The next result gives us a way to calculate the extrinsic mean and establishes its consistency of the sample mean as an estimator of the population mean if that exists.

Theorem 2.1 **(Consistency of the Extrinsic Sample Mean)** Assume the same set-up as in Definition 2.1. (a) If $\tilde{\mu} = \int_{R^k} u\tilde{Q}(du)$ is the mean of $\tilde{Q}$, then the extrinsic mean set of Q is given by $\phi^{-1}(P_{\tilde{M}}^{\tilde{\mu}})$. (b) If $\tilde{\mu}$ is a nonfocal point of R^k (relative to $\tilde{M}$), then the extrinsic sample mean μ_n (any measurable selection from the extrinsic mean set of Q_n) is a strongly consistent estimator of the extrinsic mean $\mu = \phi^{-1}(P_{\tilde{M}}^{\tilde{\mu}})$.

Proof (a) In view of the isometry $\phi : M \to \tilde{M}$, it is enough to prove that the Frechet function $\tilde{F}(u) = \int_{\tilde{M}} \|u - v\|^2 \tilde{Q}(dv)$ (on $\tilde{M}$) is minimum iff $u \in P_{\tilde{M}}\tilde{\mu}$. Now for arbitrary $u, v \in \tilde{M}$, one has

$$\|u - v\|^2 = \|u - \tilde{\mu}\|^2 + \|\tilde{\mu} - v\|^2 + 2(u - \tilde{u}) \cdot (\tilde{\mu} - v)$$

where $\cdot$ denotes Euclidean inner product. Integrate both sides wrt $\tilde{Q}(dv)$ to get

$$\tilde{F}(u) = \|u - \tilde{\mu}\|^2 + \int_{\tilde{M}} \|\tilde{\mu} - v\|^2 \tilde{Q}(dv) \tag{2.1}$$

since the integral of the product term equals $2(u - \tilde{\mu}) \cdot (\tilde{\mu} - \tilde{\mu}) = 0$. Minimization of the RHS of (2.1) is achieved precisely for $u \in P_{\tilde{M}}\tilde{\mu}$ since the second term on the right does not involve u.

(b) Since $\tilde{M}$ is closed, closed bounded subsets of $\tilde{M}$ are compact. Therefore consistency of the sample mean $\tilde{Q}_n$ follows from Theorem 1.2, if one notes that $P_{\tilde{M}}\tilde{\mu}$ is the unique minimizer of $\tilde{F}$, since $\tilde{\mu}$ is non-focal. □

Now we see examples of some Riemannian manifolds where we compute the extrinsic mean and variation.

2.1 Example1: Unit sphere S^{k-1}

Consider the inclusion map $i : S^{k-1} \to \mathbb{R}^k$, $i(x) = x$. The extrinsic mean set of a probability measure Q on S^{k-1} is then the point(set) $P_{S^{k-1}}\tilde{\mu}$ on S^{k-1} closest to $\tilde{\mu} = \int_{\mathbb{R}^k} x\tilde{Q}(dx)$, where $\tilde{Q}$ is Q regarded as a probability measure on $\mathbb{R}^k$. Note that $\tilde{\mu}$ is non-focal iff $\tilde{\mu} \neq 0$. Then $P_{S^{k-1}}\tilde{\mu} = \frac{\tilde{\mu}}{\|\tilde{\mu}\|}$, else $P_{S^{k-1}}(0) = S^{k-1}$.

From (2.1), the extrinsic variation of Q is:

$$\begin{aligned} V &= \int_{\mathbb{R}^k} \|x - \tilde{\mu}\|^2 \tilde{Q}(dx) + (\|\tilde{\mu}\| - 1)^2 \\ &= 2(1 - \|\tilde{\mu}\|) \end{aligned}$$

So that $\|\tilde{\mu}\| = 1$ iff Q is degenerate at a point.

2.2 Example2: Axial Space $\mathbb{R}P^{k-1}$

Consider the real projective space $\mathbb{R}P^{k-1}$ of all lines $(\lambda x : \lambda \in \mathbb{R} \setminus \{0\})$ through the origin in $\mathbb{R}^k, x \neq 0$. Each such line is specified by its points of intersection with the unit sphere S^{k-1}. In other words, $\mathbb{R}P^{k-1}$ may be regarded as the quotient space of S^{k-1} under the equivalence relation $u \sim v$ iff $u = -v$. The elements of $\mathbb{R}P^{k-1}$ may be represented as $[u] = \{-u, u\}(u \in S^{k-1})$.
Another representation of $\mathbb{R}P^{k-1}$ is via the **Veronese-Whitney embedding** ϕ of $\mathbb{R}P^{k-1}$ into the space of all $k \times k$ matrices identified with $\mathbb{R}^{k^2}$ by arranging the k^2 elements of a matrix as a k^2-dimensional vector, with ϕ given by

$$\phi([u]) = uu' = ((u_i u_j))_{1 \leq i,j \leq k} \quad (u = (u_1, .., u_k)' \in S^{k-1}) \tag{2.2}$$

As $\mathbb{R}^{k^2}$, the space of $k \times k$ matrices has the Euclidean distance

$$\|A - B\|^2 \equiv \sum_{1 \leq i,j \leq k} (a_{ij} - b_{ij})^2 = Trace(A - B)(A - B)' \tag{2.3}$$

If we now define the extrinsic distance d on $\mathbb{R}P^{k-1}$ as

$$d^2([u], [v]) = \|uu' - vv'\|^2 = Trace(uu' - vv')^2 \tag{2.4}$$

then $(\mathbb{R}P^{k-1}, d)$ satisfy the hypothesis of Proposition 1.1. Also $\phi(\mathbb{R}P^{k-1})$ is closed.
Let Q be a probability measure on $\mathbb{R}P^{k-1}$, and let $\tilde{\mu}$ be the mean of $\tilde{Q} = Q \circ \phi^{-1}$ considered as a probability measure on $\mathbb{R}^{k^2}$. Since $\tilde{\mu}$ is a mixture of elements of the form uu', $\tilde{\mu} \in S^+(k, \mathbb{R})$: the space of all symmetric non-negative definite $k \times k$ matrices with real elements. We need to identify the set of all nonfocal $\tilde{\mu}$ in $S^+(k, \mathbb{R})$. Let then $\tilde{\mu} \in S^+(k, \mathbb{R})$. There exists an orthogonal $k \times k$ matrix T such that $T\tilde{\mu}T' = D \equiv Diag(\lambda_1, \ldots, \lambda_k)$ where the eigen values may be taken to be ordered: $0 \leq \lambda_1 \leq \ldots \leq \lambda_k$. To find $P_{\tilde{M}}\tilde{\mu}$ with $\tilde{M} = \phi(RP^{k-1}) \equiv$ set of all matrices of the form (2.2), note first that, writing $v = Tu$

$$\begin{aligned} \|\tilde{\mu} - uu'\|^2 &\equiv Trace(\tilde{\mu} - uu')(\tilde{\mu} - uu') \\ &= Trace(T(\tilde{\mu} - uu')T')(T(\tilde{\mu} - uu')T') = Trace(D - vv')^2 \end{aligned} \tag{2.5}$$

Hence

$$
\begin{aligned}
\|\tilde{\mu} - uu'\|^2 &= \sum_{i=1}^{i=k}(\lambda_i - v_i^2)^2 + \sum_{j\neq j'}(v_i v_j)^2 \\
&= \sum_{i=1}^{k}\lambda_i^2 + \sum_{i=1}^{k} v_i^4 - 2\sum_{i=1}^{k}\lambda_i v_i^2 + (\sum_{j} v_j^2)(\sum_{j'} v_{j'}^2) - \sum_{j=1}^{k} v_j^4 \\
&= \sum_{i=1}^{k}\lambda_i^2 - 2\sum_{i=1}^{k}\lambda_i v_i^2 + 1 \qquad (2.6)
\end{aligned}
$$

The minimum is achieved if $v = (0, 0, \ldots, 0, 1) = e_k$ (since λ_k is the largest eigenvalue of μ). Since $\mu \equiv T'v = T'e_k$ is the eigenvector of $\tilde{\mu}$ having the eigenvalue λ_k, the minimum distance between $\tilde{\mu}$ and $\tilde{M}$ is attained by $[\mu\mu']$ where μ is a unit vector in the eigenspace of the largest eigenvalue of $\tilde{\mu}$. Thus $\tilde{\mu}$ is **nonfocal** iff its largest eigenvalue is **simple**, i.e., if the eigenspace corresponding to the largest eigenvalue is one dimensional. In that case the extrinsic mean of Q is $[\mu]$. Therefore, we have the following consequence of Theorem 2.1.

Proposition 2.2 Assume that the largest eigenvalue of $\tilde{\mu} = \int uu'Q(d[u])$ is simple. Let μ_n denote a eigenvector of $\frac{1}{n}\sum_{i=1}^{n} X_i X_i{}'$ (where $X_i, 1 \leq i \leq n$, are iid such that $[X_1]$ has distribution Q), having the largest eigenvalue. Then $[\mu_n]$ is a strongly consistent estimator of the Extrinsic mean $[\mu]$.

Also from (2.1) and (2.6), we get the Extrinsic variation of Q to be

$$
\begin{aligned}
V &= E\|X_1X_1' - \mu\mu'\|^2 \\
&= E\|X_1X_1' - \tilde{\mu}\|^2 + \sum_{i=1}^{k}\lambda_i^2 - 2\lambda_k + 1 \\
&= E\,Trace(X_1X_1' - \tilde{\mu})^2 + \sum_{i=1}^{k}\lambda_i^2 - 2\lambda_k + 1 \\
&= 1 - Trace(\tilde{\mu}^2) + \sum_{i=1}^{k}\lambda_i^2 - 2\lambda_k + 1 \\
&= 2(1 - \lambda_k)
\end{aligned}
$$

2.3 Example 3: Planer Shape Space of k-ads

Consider a set of k points on the plane, e.g., k locations on a skull projected on a plane, not all points being the same. We will assume $k > 2$ and refer to such a set as a k-ad (or a set of k landmarks).
For convenience we will denote a k-ad by k complex numbers ($z_j = x_j + iy_j, 1 \leq j \leq k$), i.e., we will represent k-ads on a complex plane.
By the shape of a k-ad $z = (z_1, z_2, \ldots, z_k)$, we mean the equivalence class, or orbit of z under translation, rotation and scaling.
To remove translation, one may substract $\langle \mathbf{z} \rangle \equiv (\langle z \rangle, \langle z \rangle, \ldots, \langle z \rangle)$ ($\langle z \rangle = \frac{1}{k}\sum_{j=1}^{k} z_j$) from z to get $z - \langle \mathbf{z} \rangle$.
Rotation of the k-ad by an angle θ and scaling (by a factor $r > 0$) are achieved by multiplying $z - \langle \mathbf{z} \rangle$ by the complex number $\lambda = r \exp i\theta$.
Hence one may represent the shape of the k-ad as the complex line passing through $z - \langle \mathbf{z} \rangle$, namely, $\{\lambda(z - \langle \mathbf{z} \rangle) \colon \lambda \in \mathbb{C} \setminus \{0\}\}$.
Thus the space of k-ads is the set of all complex lines on the (complex(k-1)-dimensional) hyperplane, $H^{k-1} = \{w \in C^k \setminus \{0\} \colon \sum_1^k w_j = 0\}$.
So the shape space Σ_2^k of planer k-ads has the structure of the complex projective space $\mathbb{C}P^{k-2}$: the space of all complex lines through the origin in $\mathbb{C}^{k-1}$. As in the case of $\mathbb{C}P^{k-2}$, it is convenient to represent the element of Σ_2^k corresponding to a k-ad $\mathbf{z}$ by the curve $\gamma(z) = [z] = \{e^{i\theta}\frac{(z-\langle \mathbf{z} \rangle)}{\|z-\langle \mathbf{z} \rangle\|} \colon 0 \leq \theta < 2\pi\}$ on the unit sphere in $H^{k-1} \approx \mathbb{C}^{k-1}$.
Denote by u the quantity $\frac{(z-\langle \mathbf{z} \rangle)}{\|z-\langle \mathbf{z} \rangle\|}$. That is called the preshape of the shape of z. Then the **Veronese-Whitney embedding** of Σ_2^k is given by

$$\begin{aligned} &\phi \colon \Sigma_2^k \to \mathbb{C}^{k^2}, \\ &\phi([z]) = uu^* \;\; (u = (u_1, \ldots, u_k)' \in H^{k-1}, \|u\| = 1) \\ &\;= ((u_i \bar{u}_j))_{1 \leq i,j \leq k} \end{aligned} \tag{2.7}$$

The shape of z, $[z] = \{e^{i\theta}u \colon 0 \leq \theta < 2\pi\}$ is the orbit of the vector u under rotation. Note that if $v_1, v_2 \in [z]$, then $\phi([v_1]) = \phi([v_2]) = \phi(\frac{(z-\langle \mathbf{z} \rangle)}{\|z-\langle \mathbf{z} \rangle\|})$.
Define the extrinsic distance d on Σ_2^k by that induced from this embedding, namely,

$$d^2([z],[w]) = \|uu^* - vv^*\|^2 \;\; ,u \doteq \frac{z - \langle \mathbf{z} \rangle}{\|z - \langle \mathbf{z} \rangle\|} \; ,v \doteq \frac{w - \langle \mathbf{w} \rangle}{\|w - \langle \mathbf{w} \rangle\|} \tag{2.8}$$

where for arbitrary $k \times k$ complex matrices A, B

$$\|A - B\|^2 = \sum_{j,j'} |a_{jj'} - b_{jj'}\|^2 = Trace(A - B)(A - B)^* \qquad (2.9)$$

is just the squared euclidean distance between A and B regarded as elements of $\mathbb{C}^{k^2}$ (or, $\mathbb{R}^{2k^2}$).

Since the matrices uu^*, vv^* in (2.8) are Hermitian, one notes that the image $\phi(\Sigma_2^k)$ of Σ_2^k is a closed subset of $\mathbb{C}^{k^2}$ and the "conjugate-transpose" symbol * may be dropped from (2.9) in computing distances in $\phi(\Sigma_2^k)$.

Let Q be a probability measure on the shape space Σ_2^k, and let μ_0 denote the mean vector of $Q_0 \doteq Q \circ \phi^{-1}$, regarded as a probability measure on $\mathbb{C}^{k^2}$ (or, $\mathbb{R}^{2k^2}$). Note that μ_0 belongs to the convex hull of $\phi(\Sigma_2^k)$ and in particular, is an element of H^{k-1}. Let T be a (complex) orthogonal $k \times k$ matrix such that $T\mu_0 T^* = D = Diag(\lambda_1, \lambda_2, \ldots, \lambda_k)$, where $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k$ are the eigenvalues of μ_0. Then, writing $v = Tu$ with u as in (2.8),

$$\begin{aligned}\|uu^* - \mu_0\|^2 &= \|vv^* - D\|^2 = \sum_{j=1}^{k}(|v_j|^2 - \lambda_j)^2 + \sum_{j \neq j'} |v_j \overline{v}_{j'}|^2 \\ &= \sum {\lambda_j}^2 + \sum_{j=1}^{k} |v_j|^4 - 2\sum_{j=1}^{k} \lambda_j |v_j|^2 + \sum_{j=1}^{k} |v_j|^2 . \sum_{j'=1}^{k} |v_{j'}|^2 - \sum_{j=1}^{k} |v_j|^4 \\ &= \sum {\lambda_j}^2 + 1 - 2\sum_{j=1}^{k} \lambda_j |v_j|^2\end{aligned}$$

which is minimized (on $\phi(\Sigma_2^k)$) by taking $v = e_k = (0, \ldots, 0, 1)'$, i.e., $u = T^* e_k$ - a unit eigenvector having the largest eigenvalue λ_k of μ_0. It follows that the Extrinsic mean μ, say, of Q is unique iff the eigenspace for the largest eigenvalue of μ is (complex) one dimensional, and then $\mu = [w]$, $w(\neq 0) \in$ the eigenspace of the largest eigenvalue of μ_0. It follows from Theorem 2.1 that any measurable selection from the sample extrinsic mean set is a consistent estimator of μ iff the largest eigenvalue of μ_0 is simple, i.e. it has an eigenspace of complex dimension one.

By similar analysis as in the Real Projective space, one can show that the Extrinsic Variation of Q has the expression

$$V = 2(1 - \lambda_k)$$

The distance d on Σ_2^k (see (2.8)) may be expressed as

$$\begin{aligned} d^2([z],[w]) &\equiv \|uu^* - vv^*\|^2 = \sum_{j,j'} |u_j\overline{u}_{j'} - v_j\overline{v}_{j'}|^2 \\ &= \sum_{j,j'} |u_j\overline{u}_{j'}|^2 + \sum_{j,j'} |v_j\overline{v}_{j'}|^2 - 2\sum_{j,j'} |u_j\overline{u}_{j'}v_j\overline{v}_{j'}| \\ &= \sum_{j} |u_j|^2 \sum_{j'} |\overline{u}_{j'}|^2 + \sum_{j} |v_j|^2 \sum_{j'} |\overline{v}_{j'}|^2 - 2(u^*v)(v^*u) \\ &= 2 - 2(u^*v)(v^*v) \\ &= 2(1 - |u^*v|^2) \end{aligned} \tag{2.10}$$

with u and v as in (2.8). This is the so-called **Full Procrustes distance** for Σ_2^k (See Dryden and Mardia [6]).
Let $[z]$ and $[w]$ be two shapes and let u and v be their preshapes. Then the **Procrustes coordinates** of v onto u is defined as

$$v^P = (\hat{a} + i\hat{b})1_k + \hat{\beta}e^{i\hat{\theta}}v.$$

where (β, θ, a, b) are chosen to minimize

$$D^2(u,v) = \|u - \beta e^{i\theta}v - (a+ib)1_k\|^2$$

Then one gets

$$\begin{aligned} \hat{a} &= \hat{b} = 0 \\ \hat{\beta}e^{i\hat{\theta}} &= v^*u \\ \text{So } v^P &= (v^*u)v \\ \text{and } D^2(u,v) &= (1 - |v^*u|^2) = \frac{1}{2}d^2([z],[w]) \end{aligned}$$

Here 1_k denotes the vector of ones of length k, that is $1_k = (1, 1, \ldots, 1)_k'$.

As a numerical example, consider 8 locations on a gorilla skull projected on a plane. There are 30 female and 29 male samples. [Source: Statistical Shape Analysis - Dryden & Mardia]. Figures 1 and 2 are the plot of the Procrustes coordinates of the female and male samples onto their extrinsic sample means respectively.

3 Asymptotic Distribution of the Extrinsic Sample Mean

One can use Theorem 1.3 to get the asymptotic distribution of the sample mean. However expressions for the parameters Λ and Σ are not easy to get. Here we devise another way to deduce the asymptotic distribution.
We are in the same set up as in the start of Section 2. ϕ is an embedding of M into $\mathbb{R}^k$. The mean μ of the image $\tilde{Q} = Q \circ \phi^{-1}$ is a non-focal point of $\mathbb{R}^k$, so that the projection $P(\mu)$ of μ on $\phi(M)$ is unique, and the extrinsic mean of Q is $\mu_E = \phi^{-1}P(\mu)$.

Let $\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$ denote the sample mean of $Y_j = \phi(X_j)$, where $X_1, \ldots, X_n$ is a random sample from Q. The extrinsic sample mean is $\phi^{-1}(P(\bar{Y}))$, where $P(\bar{Y})$ is the projection of $\bar{Y}$ on $\phi(M)$. In a neighborhood of a nonfocal point such as μ, $P(.)$ is smooth. Write

$$\sqrt{n}[P(\bar{Y}) - P(\mu)] = \sqrt{n}(d_\mu P)(\bar{Y} - \mu) + o_P(1) = (d_\mu P)(\sqrt{n}(\bar{Y} - \mu)) + o_P(1) \tag{3.1}$$

where $d_\mu P$ is the differential (map) of the projection $P(.)$, which takes vectors in the tangent space of $\mathbb{R}^k$ at μ to tangent vectors of $\phi(M)$ at $P(\mu)$. Let $f_1, f_2, \ldots, f_d$ be an orthonormal basis of $T_{P(\mu)}\phi(M)$ and $e_1, e_2, \ldots, e_k$ be an orthonormal basis (frame) of $\mathbb{R}^k$. One has

$$\begin{aligned} \sqrt{n}(\bar{Y} - \mu) &= \sum_{j=1}^{k} \sqrt{n}(\bar{Y} - \mu)_j e_j, \\ d_\mu P(\sqrt{n}(\bar{Y} - \mu)) &= \sum_{j=1}^{k} \sqrt{n}(\bar{Y} - \mu)_j d_\mu P(e_j) \\ &= \sum_{j=1}^{k} \sqrt{n}(\bar{Y} - \mu)_j \sum_{r=1}^{d} a_{jr}(\mu) f_r \\ &= \sum_{r=1}^{d} [\sum_{j=1}^{k} a_{jr}(\mu)\sqrt{n}(\bar{Y} - \mu)_j] f_r \qquad (3.2) \end{aligned}$$

where $a_{jr}(\mu) = d_\mu P(e_j) \cdot f_r$. Hence, $\sqrt{n}[P(\bar{Y}) - P(\mu)]$ has an asymptotic Gaussian distribution on the tangent space of $\phi(M)$ at $P(\mu)$, with mean vector zero and a dispersion matrix (wrt the basis vector $\{f_r : 1 \le r \le d\}$)

$$\Sigma = A'VA \ \ [A = A(\mu) = ((a_{jr}(\mu)))_{1\le j\le k, 1\le r\le d}] \tag{3.3}$$

V being the $k \times k$ covariance matrix of $\tilde{Q}$ (wrt the basis $\{e_j : 1 \le j \le k\}$). In matrix notation,

$$\sqrt{n}A'(\bar{Y} - \mu) \xrightarrow{\mathcal{L}} N(0, \Sigma) \ \text{ as } n \to \infty \tag{3.4}$$

This implies, writing $\mathcal{X}_d^2$ for the chisquare distribution with d degrees of freedom,

$$\begin{aligned} &n(\bar{Y} - \mu)' A\Sigma^{-1}A'(\bar{Y} - \mu) \longrightarrow \mathcal{X}_d^2, \\ &n(\bar{Y} - \mu)' A\hat{\Sigma}^{-1}A'(\bar{Y} - \mu) \longrightarrow \mathcal{X}_d^2 \text{ as } n \to \infty \end{aligned} \tag{3.5}$$

Here

$$\begin{aligned} \hat{\Sigma} &= (A(\mu))'\hat{V}A(\mu), \\ \hat{V} &= ((\frac{1}{n}\sum_{i=1}^{n}(Y_{ij} - \mu_j)(Y_{ij'} - \mu_{j'})))_{j,j'=1}^{k} \end{aligned} \tag{3.6}$$

A confidence region for $P(\mu)$ with asymptotic confidence level $1 - \alpha$ is then given by

$$\{P(\mu) : n(\bar{Y} - \mu)' A\hat{\Sigma}^{-1}A'(\bar{Y} - \mu) \le \mathcal{X}_d^2(1 - \alpha)\} \tag{3.7}$$

Note that $A = A(\mu)$ depends on μ. The bootstrapped version of the statistic in (3.7) is

$$\begin{aligned} U^* &= n(\bar{Y}^* - \bar{Y})' A(\bar{Y})(A(\bar{Y})'\hat{V}^* A(\bar{Y}))^{-1}A'(\bar{Y})(\bar{Y}^* - \bar{Y}), \\ \hat{V}^* &= ((\frac{1}{n}\sum_{i=1}^{n}(Y_{ij}^* - \bar{Y}_j)(Y_{ij'}^* - \bar{Y}_{j'})))_{j,j'=1}^{k} \end{aligned} \tag{3.8}$$

The corresponding bootstrapped confidence region is given by

$$\{P(\mu) : n(\bar{Y} - \mu)' A\hat{\Sigma}^{-1}A'(\bar{Y} - \mu) \le c^*_{(1-\alpha)}\} \tag{3.9}$$

where $c^*_{(1-\alpha)}$ is the upper $(1-\alpha)$-quantile of the bootstrapped values U^*. An alternative to (2.17), simpler to compute, is given by

$$\begin{aligned} &\{P(\mu) : n(\bar{Y} - \mu)' A(\bar{Y})\hat{\hat{\Sigma}}^{-1}A(\bar{Y})'(\bar{Y} - \mu) \le \mathcal{X}_d^2(1 - \alpha)\}, \\ &\hat{\hat{\Sigma}} = A'(\bar{Y})\hat{V}A(\bar{Y}) \end{aligned} \tag{3.10}$$

But then the corresponding bootstrapped version becomes even more computation intensive than (3.8).

3.1 Asymptotic distribution of the mean shape

As an application, let us find the asymptotic distribution of the sample extrinsic mean shape of a sample of size n from the planer shape space.
$M = \Sigma_2^k$ can be embedded into $S(k, \mathbb{C})$, the space of all $k \times k$ complex self adjoint matrices, via the map ϕ in (2.7).
We consider $S(k, \mathbb{C})$ as a linear subspace of $\mathbb{C}^{k^2}$ (over $\mathbb{R}$) and as such a regular submanifold of $\mathbb{C}^{k^2}$ embedded by the inclusion map, and inheriting the metric:

$$\langle A, B \rangle = Re\ Trace(A\bar{B}')$$

The dimension of $S(k, \mathbb{C})$ is k^2. An orthonormal basis for $S(k, \mathbb{C})$ is given by $\{v_b^a : 1 \leq a \leq b \leq k\}$ and $\{w_b^a : 1 \leq a < b \leq k\}$:

$$\begin{aligned} v_b^a &= \frac{1}{\sqrt{2}}(e_a e_b^t + e_b e_a^t), \quad a < b \\ &= e_a e_a^t, \quad a = b \\ w_b^a &= \frac{i}{\sqrt{2}}(e_a e_b^t - e_b e_a^t), \quad a < b. \end{aligned}$$

where $\{e_a : 1 \leq a \leq k\}$ is the standard canonical basis for $\mathbb{R}^k$.
We also take $\{v_b^a : 1 \leq a \leq b \leq k\}$ and $\{w_b^a : 1 \leq a < b \leq k\}$ as the orthogonal frame for $TS(k, \mathbb{C}) \equiv S(k, \mathbb{C})$. Note that for all $U \in O(k)$ $(UU^* = U^*U = I)$, $\{Uv_b^aU^* : 1 \leq a \leq b \leq k\}$, $\{Uw_b^aU^* : 1 \leq a < b \leq k\}$ is also an orthogonal frame for $S(k, \mathbb{C})$.

Assume that the mean μ of $\tilde{Q} = Q \circ \phi^{-1}$ has its largest eigen value simple. Then from (3.1), one has

$$\sqrt{n}[P(\bar{Y}) - P(\mu)] = d_\mu P(\sqrt{n}(\bar{Y} - \mu)) + o_P(1) \tag{3.11}$$

Here we view $d_\mu P : S(k, \mathbb{C}) \to T_{P(\mu)}\phi(\Sigma_2^k)$. Choose $U \in O(k)$ such that $U^*\mu U = D \equiv Diag(\lambda_1, \ldots, \lambda_k)$, $\lambda_1 \leq \ldots \leq \lambda_{k-1} < \lambda_k$ being the eigenvalues of μ.

Choose a basis $\{Uv_b^aU^*, Uw_b^aU^*\}$ for $S(k, \mathbb{C})$. Then one can show that

$$d_\mu P(Uv_b^aU^*) = \begin{cases} 0 & \text{if } 1 \leq a \leq b < k, \text{ a = b = k}, \\ (\lambda_k - \lambda_a)^{-1} Uv_k^aU^*, & 1 \leq a < k, b = k. \end{cases}$$

and

$$d_\mu P(Uw_b^aU^*) = \begin{cases} 0 & \text{if } 1 \leq a < b < k \\ (\lambda_k - \lambda_a)^{-1} Uw_k^aU^*, & 1 \leq a < k, b = k. \end{cases} \tag{3.12}$$

Write

$$\begin{aligned} \sqrt{n}(\bar{Y} - \mu) = & \sum\sum_{1 \leq a \leq b \leq k} < \sqrt{n}(\bar{Y} - \mu), Uv_b^aU^* > Uv_b^aU^* \\ & + \sum\sum_{1 \leq a < b \leq k} < \sqrt{n}(\bar{Y} - \mu), Uw_b^aU^* > Uw_b^aU^* \end{aligned} \tag{3.13}$$

Since $\bar{Y}\mathbf{1}_k = \mu\mathbf{1}_k = 0$, $\lambda_1 = 0$ and $U(:, 1) = \alpha\mathbf{1}_k$, $|\alpha| = \frac{1}{\sqrt{k}}$. It is easy to check that $\langle\sqrt{n}(\bar{Y} - \mu), Uv_b^1U^*\rangle = \langle\sqrt{n}(\bar{Y} - \mu), Uw_b^1U^*\rangle = 0$.
So from (3.12) and (3.13),

$$\begin{aligned} & d_\mu P(\sqrt{n}(\bar{Y} - \mu)) \\ = & \sum_{a=2}^{k-1} < \sqrt{n}(\bar{Y} - \mu), Uv_k^aU^* > (\lambda_k - \lambda_a)^{-1} Uv_k^aU^* \\ + & \sum_{a=2}^{k-1} < \sqrt{n}(\bar{Y} - \mu), Uw_k^aU^* > (\lambda_k - \lambda_a)^{-1} Uw_k^aU^* \end{aligned} \tag{3.14}$$

From (3.11) and (3.14), we see that $\sqrt{n}(P(\bar{Y}) - P(\mu))$ has an asymptotic Gaussian distribution on a subspace of $S(k, \mathbb{C})$ with asymptotic coordinates $T_n(\mu) = (\langle\sqrt{n}(\bar{Y} - \mu), Uv_k^aU^*\rangle_{a=2}^{k-1}, \langle\sqrt{n}(\bar{Y} - \mu), Uw_k^aU^*\rangle_{a=2}^{k-1})$ wrt the basis vector $\{(\lambda_k - \lambda_a)^{-1}Uv_k^aU^*, (\lambda_k - \lambda_a)^{-1}Uw_k^aU^*\}_{a=2}^{k-1}$.
Then

$$T_n(\mu)'\Sigma(\mu)^{-1}T_n(\mu) \longrightarrow \mathcal{X}_{2k-4}^2$$

Write $\tilde{U} = U(:, 2 : k - 1)$. Then

$$\begin{aligned} & T_n(\mu) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \tilde{T}_j(\mu) \text{ where} \\ & \tilde{T}_j(\mu)' = \sqrt{2}(Re(\tilde{U}^*Y_jU(:, k))', \ Im(\tilde{U}^*Y_jU(:, k))') \\ & \Sigma(\mu) = E\tilde{T}_j(\mu)\tilde{T}_j(\mu)' \end{aligned}$$

One can estimate $\Sigma(\mu)$ by $\hat{\Sigma}(\mu)$: the sample covariance matrix, or $\hat{\Sigma}(\bar{Y})$: the sample covariance matrix, with U replaced by the eigen vectors of $\bar{Y}$, as in (3.6) and (3.10).
This will give confidence sets for $P(\mu)$ as is (3.7) and (3.10).

3.2 A two sample testing problem

Let Q_1 and Q_2 be two probability measures on the shape space Σ_2^k, and let μ_1 and μ_2 denote the mean vectors of $Q_1 \circ \phi^{-1}$ and $Q_2 \circ \phi^{-1}$ respectively. Suppose $[x_1], \ldots, [x_n]$ and $[y_1], \ldots, [y_m]$ are iid random samples from Q_1 and Q_2 respectively. Let $X_i = \phi([x_i])$, $Y_i = \phi([y_i])$ be their images onto $\phi(\Sigma_2^k)$ which are random samples from $Q_1 \circ \phi^{-1}$ and $Q_2 \circ \phi^{-1}$ respectively. Suppose we are to test if the extrinsic means of Q_1 and Q_2 are equal, i.e.

$$H_0 : P\mu_1 = P\mu_2$$

We assume that both μ_1 and μ_2 have simple largest eigen values. Then under H_0, the corresponding eigen vectors differ by a rotation.
Choose $\mu \in S(k, \mathbb{C})$ with same projection as μ_1 and μ_2. Suppose $\mu = U\Lambda U^*$, where $\Lambda = Diag(\lambda_1 \leq \lambda_2 \leq \ldots < \lambda_k)$ are its eigen values and $U = [U_1, U_2, \ldots, U_k]$ are the corresponding eigen vectors. Under H_0, $P\mu_1 = P\mu_2 = U_k U_k^*$.
From Section 3.1,

$$\begin{aligned} d_\mu P(\bar{X} - \mu) &= \sum_{a=2}^{k-1} \sqrt{2} Re(U_a^* \bar{X} U_k)(\lambda_k - \lambda_a)^{-1} U v_k^a U^* + \sum_{a=2}^{k-1} \sqrt{2} Im(U_a^* \bar{X} U_k)(\lambda_k - \lambda_a)^{-1} U w_k^a U^* \\ &= \sum_{a=2}^{k-1} (\lambda_k - \lambda_a)^{-1} (U_a^* \bar{X} U_k) U_a U_k^* + \sum_{a=2}^{k-1} (\lambda_k - \lambda_a)^{-1} (U_k^* \bar{X} U_a) U_k U_a^* \end{aligned} \tag{3.15}$$

and

$$\begin{aligned}d_\mu P(\bar{Y}-\mu) &= \sum_{a=2}^{k-1}\sqrt{2}Re(U_a^*\bar{Y}U_k)(\lambda_k-\lambda_a)^{-1}Uv_k^aU^* + \sum_{a=2}^{k-1}\sqrt{2}Im(U_a^*\bar{Y}U_k)(\lambda_k-\lambda_a)^{-1}Uw_k^aU^* \\ &= \sum_{a=2}^{k-1}(\lambda_k-\lambda_a)^{-1}(U_a^*\bar{Y}U_k)U_aU_k^* + \sum_{a=2}^{k-1}(\lambda_k-\lambda_a)^{-1}(U_k^*\bar{Y}U_a)U_kU_a^*\end{aligned} \tag{3.16}$$

Let $T_n(\mu) = (Re(U_a^*\bar{X}U_k),\ Im(U_a^*\bar{X}U_k))_{a=2}^{k-1}$ and
$S_m(\mu) = (Re(U_a^*\bar{Y}U_k),\ Im(U_a^*\bar{Y}U_k))_{a=2}^{k-1}$

Then under H_0, $T_n(\mu)$ and $S_m(\mu)$ have mean zero, and as $n, m \to \infty$,

$$\begin{aligned}\sqrt{n}T_n(\mu) &\xrightarrow{\mathcal{L}} N(0, \Sigma_1(\mu)) \\ \text{and } \sqrt{m}S_m(\mu) &\xrightarrow{\mathcal{L}} N(0, \Sigma_2(\mu))\end{aligned}$$

Suppose $\frac{n}{m+n} \to p$, $\frac{m}{m+n} \to q$, for some $p, q > 0$; $p + q = 1$. Then

$$\sqrt{n+m}(T_n(\mu) - S_m(\mu)) = \sqrt{n+m}(Re(U_a^*(\bar{X}-\bar{Y})U_k),\ Im(U_a^*(\bar{X}-\bar{Y})U_k)_{a=2}^{k-1}$$
$$\xrightarrow{\mathcal{L}} N_{2k-4}(0, \frac{1}{p}\Sigma_1 + \frac{1}{q}\Sigma_2)$$

So $(n+m)(T_n(\mu) - S_m(\mu))'(\frac{1}{p}\Sigma_1(\mu) + \frac{1}{q}\Sigma_2(\mu))^{-1}(T_n(\mu) - S_m(\mu)) \xrightarrow{\mathcal{L}} \mathcal{X}_{2k-4}^2$ (3.17)

We can choose μ to be any positive linear combination of μ_1 and μ_2. Then under H_0, μ will have same projection on $\phi(\Sigma_2^k)$ as μ_1 and μ_2. We may take $\mu = p\mu_1 + q\mu_2$.

In practice, since μ_1 and μ_2 are unknown, so is μ. Then we may estimate μ by the pooled sample mean, $\hat{\mu} = \frac{n\bar{X}+m\bar{Y}}{m+n}$; $\Sigma_1(\mu)$ and $\Sigma_2(\mu)$ by their sample estimates $\hat{\Sigma}_1(\hat{\mu})$ and $\hat{\Sigma}_2(\hat{\mu})$, where

$$\hat{\Sigma}_1(\mu) = \frac{1}{n}X(\mu)X(\mu)' - \bar{X}(\hat{\mu})\bar{X}(\hat{\mu})'$$

$$X(\mu)_{ij} = \begin{cases} Re(U_{i+1}^* X_j U_k) & \text{if } 1 \leq i \leq k-2,\ 1 \leq j \leq n \\ Im(U_{i-k+3}^* X_j U_k) & \text{if } k-1 \leq i \leq 2k-4,\ 1 \leq j \leq n \end{cases}$$

$$\bar{X}(\hat{\mu}) = \frac{1}{n}\sum_{j=1}^{n} X(\hat{\mu})_{.j}$$

and

$$\hat{\Sigma}_2(\mu) = \frac{1}{m}Y(\mu)Y(\mu)' - \bar{Y}(\hat{\mu})\bar{Y}(\hat{\mu})'$$

$$Y(\mu)_{ij} = \begin{cases} Re(U_{i+1}^* Y_j U_k) & \text{if } 1 \leq i \leq k-2,\ 1 \leq j \leq m \\ Im(U_{i-k+3}^* Y_j U_k) & \text{if } k-1 \leq i \leq 2k-4,\ 1 \leq j \leq m \end{cases}$$

$$\bar{Y}(\hat{\mu}) = \frac{1}{m}\sum_{j=1}^{m} Y(\hat{\mu})_{.j}$$

Then the two sample test statistic in (3.17) can be estimated by

$$(\bar{X}(\hat{\mu}) - \bar{Y}(\hat{\mu}))'(\frac{1}{n}\hat{\Sigma}_1(\hat{\mu}) + \frac{1}{m}\hat{\Sigma}_2(\hat{\mu}))^{-1}(\bar{X}(\hat{\mu}) - \bar{Y}(\hat{\mu})) \tag{3.18}$$

For the skull data discussed in Section 2.3, suppose we are to test if the male and female populations have the same mean shape.
Figure 3 is the plot of the full Procrustes coordinates for the (sample) Extrinsic mean shapes of female and male skulls onto the Extrinsic mean for the pooled sample.
Value of the test statistic in (3.18) = 392.68.
P-value for the test using chi square approximation =0.
So we reject H_0 and conclude that the mean shapes are different.

4 Asymptotic distribution of Extrinsic Variation

We are in the same set up as in the start of Section 2. Suppose V is the Extrinsic variation of Q, and V_n its sample analogue. Then from Theorem

1.5,

$$\sqrt{n}(V_n - V) \xrightarrow{\mathcal{L}} N(0, \sigma^2) \tag{4.1}$$

$$\text{where } \sigma^2 = \int_M (d^2(x, \mu_E) - V)^2 Q(dx)$$

The above result can be used to construct a asymptotic level α confidence interval for the population variation which is given by :

$$(V_n - \frac{s}{\sqrt{n}} Z_{1-\frac{\alpha}{2}}, V_n + \frac{s}{\sqrt{n}} Z_{1-\frac{\alpha}{2}}) \tag{4.2}$$

where $s^2 = \frac{1}{n} \sum_{j=1}^{n} (d^2(X_j, \mu_{nE}) - V_n)^2$ is the sample estimate of σ^2, μ_{nE} is the sample extrinsic mean and $Z_{1-\frac{\alpha}{2}}$ is the upper $(1 - \frac{\alpha}{2})$ quantile for standard normal distribution.

For the gorilla skull data, 95% C.I. for the variations of females and males are:
Females: $(0.0031, 0.0046)$
Males: $(0.0034, 0.0056)$

4.1 Testing equality of Extrinsic Variations

Result (4.1) can used to construct a non parametric test for testing whether two populations on the shape space have the same spread.
We are in the set up of Section 3.2. Let V_1 and V_2 denote the variations of Q_1 and Q_2 respectively, and V_{1n} and V_{2m} denote their sample analogues. Then the null hypothesis is

$$H_0 : V_1 = V_2 = V$$

Under H_0, from (4.1)

$$\sqrt{n}(V_{1n} - V) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2) \tag{4.3}$$

$$\sqrt{m}(V_{2m} - V) \xrightarrow{\mathcal{L}} N(0, \sigma_2^2) \tag{4.4}$$

$$\text{where } \sigma_1^2 = \int_{\Sigma_2^k} (d^2([u], [\mu_{1E}]) - V)^2 Q_1(d[u])$$

$$\text{and } \sigma_2^2 = \int_{\Sigma_2^k} (d^2([u], [\mu_{2E}]) - V)^2 Q_2(d[u]).$$

Suppose $\frac{n}{m+n} \to p$, $\frac{m}{m+n} \to q$, for some $p, q > 0$; $p + q = 1$. Then from (4.3) and (4.4)

$$\sqrt{n+m}(V_{1n} - V_{2m}) \xrightarrow{\mathcal{L}} N(0, \left(\frac{\sigma_1^2}{p} + \frac{\sigma_2^2}{q}\right)) \tag{4.5}$$

$$\text{So } \frac{(V_{1n} - V_{2m})}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{4.6}$$

where $s_1^2 = \frac{1}{n}\sum_{j=1}^{n}(d^2([x_j],[\mu_{nE}]) - V_{1n})^2$ and $s_2^2 = \frac{1}{m}\sum_{j=1}^{m}(d^2([y_j],[\mu_{mE}]) - V_{2m})^2$ are the sample estimates of σ_1^2 and σ_2^2 respectively and $[\mu_{nE}]$ and $[\mu_{mE}]$ are the sample mean shapes.
For the shape space, the test statistic in (4.6) becomes

$$T_{nm} = 2\frac{(\lambda_{km} - \lambda_{kn})}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}$$

where λ_{kn} and λ_{km} are the largest eigen values of $\bar{X}$ and $\bar{Y}$ respectively. The P-value for the test is $P = P(|Z| > |T_{nm}|)$ where $Z \sim N(0,1)$. We accept H_0 for large values of P.

For the skull data, $T_{nm} = -0.923$.
P = 0.356.
So we accept H_0, that is the two populations have same average spread around their respective means.

5 Intrinsic Mean and Variation

Let (M, g) be a d-dimensional connected complete Riemannian manifold, g being the riemannian metric on M. Let the distance $d = d_g$ be the geodesic distance under g. Let Q be a probability distribution on M with finite Frechet function.

Definition 5.1 The Frechet mean set of Q under the distance d_g is called its **Intrinsic Mean** set. The Frechet Variation of Q under d_g is called its **Intrinsic Variation**.
Let $X_1, X_2, \ldots, X_n$ be iid observations on M with common distribution Q. The sample Frechet mean set is called the **sample Intrinsic Mean set** and

the sample Frechet Variation is called the **sample Intrinsic Variation**.

Let us define a few technical terms related to Riemannian manifolds which we will use intensively in the subsequent sections. For more rigorous definitions see Lee: Riemannian Manifolds [8].

1. **Geodesic** These are curves γ with zero acceleration, ie $\ddot{\gamma} = 0$. They are locally length minimizing curves.
eg. Great Circles on S^n, Straight lines in $\mathbb{R}^n$.

2. **Exponential map** For $p \in M$, $V \in T_pM$; we define $exp_pV = \gamma(1)$, where γ is a geodesic with $\gamma(0) = p$ and $\dot{\gamma}(0) = V$.

3. **Cut locus** Let γ be a geodesic starting at p, $\gamma(0) = p$. Let t_0 be the supremum of all t for which γ is length minimizing on $[0, t]$. Then $\gamma(t_0)$ is called the cut point of p along γ. The cut locus of p, $C(p)$ is the set of all cut points of p along all geodesics.
eg. $C(p) = \{-p\}$ on S^n.

4. **Convex ball**: A ball B is called convex if for any $p, q \in B$, the shortest geodesic from p to q is unique in M and lies in B.
e.g. Any ball of radius $\pi/2$ on S^n is convex.

5. **Sectional Curvature**: For a curve γ, its sectional curvature at $\gamma(t)$ is $\pm|\ddot{\gamma}(t)|$, $+$ if the curve is pointing towards N and $-$ if it is pointing away from N, where N is a chosen normal field along γ.
For a $2d$ manifold, chose a basis (X, Y) for T_pM. Then the sectional curvature at p is $\frac{Rm(X,Y,Y,X)}{|X|^2|Y|^2-<X,Y>^2}$, where Rm is the 'Riemann Curvature Tensor'.
For a d dimensional manifold, consider the $2-D$ submanifold swept out by geodesics with initial velocities lying in a $2-D$ subspace, Π of T_pM. Then the sectional curvature of that submanifold is the sectional curvature at p associated with Π, $K(\Pi)$.

The next result gives a sufficient condition for the existence (uniqueness) of Intrinsic Mean.

Proposition 5.1 Suppose all sectional curvatures on M are bounded above by some $C \geq 0$. Suppose the support of Q is contained in a convex ball of radius r, $B(p,r)$ (wrt d_g) where

$$r = \begin{cases} \infty & \text{if } C = 0 \\ \frac{\pi}{4\sqrt{C}} & \text{if } C > 0 \end{cases}$$

Then the Frechet function, F of Q is strictly convex on $B(p,r)$, has a unique global minima, which is attained in $B(p,r)$ and is the unique local minima of F on $B(p,2r)$. Hence the intrinsic mean of Q exists (as a unique minimizer of F) and lies in $B(p,r)$.

Proof See Theorem 1.2, KARCHER [3] and Theorem 1, LE [4]. $\square$

In case the population intrinsic mean exists, from Theorem 1.2, the sample mean is a consistent estimator of the true mean. Now we deduce the asymptotic distribution of the sample mean after proper translation and scaling.

6 Asymptotic distribution of the sample Intrinsic mean

One can use Theorem 1.3 to get the asymptotic distribution of the sample intrinsic mean. For that we need to verify assumptions A1 to A7. The next result gives sufficient conditions for those assumptions to hold.

Proposition 6.1 Suppose the support of Q is contained in a convex geodesic ball $B(p,r)$ with center p and radius r, r as in Proposition 5.1; which is disjoint from the cutlocus of p, $C(p)$. Let $\phi = Exp_p^{-1} : B(p,r) \longrightarrow T_pM(\approx \mathbb{R}^d)$. Define $h(x,y) = d_g^2(\phi^{-1}x, \phi^{-1}y)$; $x, y \in \mathbb{R}^d$. Let $((D_r h))_{r=1}^d$ and $((D_r D_s h))_{r,s=1}^d$ be the matrix of first and second order derivatives of $y \mapsto h(x,y)$. Let $Y_j = \phi(X_j); j = 1, \ldots, n$; $X_1, \ldots, X_n$ being iid observations from Q. Let $\mu = \phi(\mu_I)$, μ_I being the intrinsic mean of Q. Let $\mu_n = \phi(\mu_{n,I})$, $\mu_{n,I}$ being a measurable selection from the sample mean set of X_j's. Define $\Lambda = E((D_r D_s h(Y_1, \mu)))_{r,s=1}^d$; $\Sigma = Cov((D_r h(Y_1, \mu)))_{r=1}^d$. Then Λ and Σ are positive definite and

$$\sqrt{n}(\mu_n - \mu) \xrightarrow{\mathcal{L}} N(0, \Lambda^{-1}\Sigma\Lambda^{-1})$$

Proof See Theorem 2.2, Bhattacharya, R. and Patrangenaru, V. [2]. $\square$

Proposition 6.1 can be used to construct an asymptotic $1-\alpha$ confidence set for μ_I which is given by

$$\{\mu_I : n(\mu_n - \mu)^t(\hat{\Lambda}^{-1}\hat{\Sigma}\hat{\Lambda}^{-1})^{-1}(\mu_n - \mu) \leq \mathcal{X}_d^2(1-\alpha)\}$$

where $\mathcal{X}_d^2(1-\alpha)$ is the upper $(1-\alpha)^{\text{th}}$ quantile of the chi-squared distribution with d degrees of freedom and

$$\hat{\Lambda} = \frac{1}{n}\sum_{i=1}^{n} D_r D_s h(Y_i, \mu_n)$$
$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} D_r h(Y_i, \mu_n) D_s h(Y_i, \mu_n)$$

are the sample estimates of Λ and Σ respectively.

Proposition 6.1 uses normal coordinates around some point p to get the asymptotic distribution of the sample mean. The natural candidate for p is the intrinsic mean of Q, μ_I. Then we have expressions for the asymptotic parameters Λ and Σ, as the next result shows.

Theorem 6.2 Suppose Q has an intrinsic mean μ_I and satisfies the following assumption:
A. For any geodesic γ, $\gamma(0) = \mu_I$; there exists $s(\mu_I) > 0$ such that the cut-locus of $\gamma\mid_{[0,s(\mu_I)]}$, $C(\gamma\mid_{[0,s(\mu_I)]})$ has Q measure 0.
This is satisfied in particular if the support of Q is contained in a convex open ball of radius r as in Proposition 5.1.
Let $Y_j = exp_{\mu_I}^{-1} X_j = (Y_j^1, \ldots, Y_j^d)$ be the normal coordinates of the sample around μ_I; Σ and Λ as defined in Proposition 6.1. Then we have the following expressions:

1. $D_r h(Y_j, \mu) = -2Y_j^r$
2. $E(D_r h(Y_1, \mu)) = 0$
3. $\Sigma_{rs} = 4Cov(Y_1^r, Y_1^s)$

If M has constant sectional curvature C, then

$$4.\ \Lambda_{rs} = 2E((\left(\frac{1-f|Y_1|}{|Y_1|^2}\right) Y_1^r Y_1^s + (f|Y_1|)\delta_{rs})),$$
$$|Y_1| = \sqrt{(Y_1^1)^2 + (Y_1^2)^2 + \dots (Y_1^d)^2}$$

where

$$f(x) = \begin{cases} 1 & \text{if } C = 0 \\ \sqrt{C}x\frac{cos(\sqrt{C}x)}{sin(\sqrt{C}x)} & \text{if } C > 0 \\ \sqrt{-C}x\frac{cosh(\sqrt{-C}x)}{sinh(\sqrt{-C}x)} & \text{if } C < 0 \end{cases} \tag{6.1}$$

Proof Let $\gamma(s)$ be a geodesic, $\gamma(0) = \mu_I$, $s < s(\mu_I)$. For $m \in M \setminus C(\gamma(s))$, define $c_s(t) = exp_m(texp_m^{-1}\gamma(s))$. For every $s < s(\mu_I)$, $c(s,.)$ is a geodesic joining m and $\gamma(s)$, so that $c(.,.)$ is a variation of γ through geodesics. Also since $m \notin C(\gamma(s))$, $c(s,.)$ is the length minimizing curve joining m and $\gamma(s)$. Let $T = D_t c(s,t), S = D_s c(s,t)$. Then $S(s,.)$ is a family of Jacobi fields along $c(s,.)$. Since $c(s,0) = m$, $S(s,0) = 0$, and since $c(s,1) = \gamma(s)$, $S(s,1) = \dot{\gamma}(s)$. $< T, T >= d(\gamma(s), m)^2$ is independent of t, $D_t T = 0$. Then

$$\begin{aligned} F(\gamma(s)) &= \int_{M\setminus C(\gamma|_{[0,s(\mu_I)]})} d(\gamma(s), m)^2 Q(dm) \\ &= \int \langle T, T\rangle Q(dm) \\ &= \int (\int_0^1 \langle T, T\rangle dt) Q(dm) && (6.2) \\ \Rightarrow \frac{d}{ds} F(\gamma(s)) &= \int (\int_0^1 2\langle D_s T, T\rangle dt) Q(dm) \\ &= 2\int (\int_0^1 \frac{d}{dt}\langle T, S\rangle dt) Q(dm) \\ &= 2\int \langle T(s,1), S(s,1)\rangle Q(dm) && (6.3) \end{aligned}$$

$S(s,1) = \dot{\gamma}(s)$ is independent of m. So

$$\frac{d}{ds} F(\gamma(s)) = 2\langle \int_M T(s,1) Q(dm), \dot{\gamma}(s)\rangle \tag{6.4}$$

Since μ_I is a local minima for F,

$$\int_M T(0,1)Q(dm) = 0.$$
$$T(0,1) = -exp_{\mu_I}^{-1}(m)$$
$$\Rightarrow \int_M exp_{\mu_I}^{-1}(m)Q(dm) = 0 \tag{6.5}$$

(6.4) and (6.5) prove 1 and 2. 3 follows from 1.
To find Λ, we need to find the second order derivatives of F. Now

$$\begin{aligned} \frac{d^2}{ds^2}F(\gamma(s)) &= 2\int < D_sT(s,1), S(s,1)\rangle Q(dm) \\ &= 2\int \langle D_tS(s,1), S(s,1)\rangle Q(dm) \end{aligned} \tag{6.6}$$

Let $J_s(t) = S(s,t)$. Then J_s is a Jacobi field along $c(s,.)$ with $J_s(0) = 0$, $J_s(1) = \dot{\gamma}(s)$ is independent of $m \in M$. Then

$$\frac{d^2}{ds^2}F(\gamma(s)) = 2\int \langle D_tJ_s(1), J_s(1)\rangle Q(dm) \tag{6.7}$$

Suppose M has constant sectional curvature C. Let $J_s^{\perp}$ and J_s^{-} be the normal and tangential components of J_s. It can be shown that

$$\langle J_s^{\perp}(1), D_tJ_s^{\perp}(1)\rangle = f(|\dot{c}(s,.)|)|J_s^{\perp}(1)|^2 \tag{6.8}$$

, f as in (6.1).
Write $J_s^{-}(t) = \lambda t\dot{c}(s,t)$ where $\lambda = \frac{\langle J_s,\dot{c}(s,.)\rangle}{|\dot{c}(s,.)|^2}$ is independent of t. Then

$$D_t(J_s^{-})(1) = (D_tJ_s)^{-}(1) = J_s^{-}(1) = \frac{\langle J_s, \dot{c}(s,1)\rangle}{|\dot{c}(s,1)|^2}\dot{c}(s,1) \tag{6.9}$$

$$\begin{aligned} \text{So } D_t|J_s^{-}|^2(1) &= 2\lambda^2|\dot{c}(s,1)|^2 = 2\frac{\langle J_s, \dot{c}(s,1)\rangle^2}{|\dot{c}(s,.)|^2} \\ &= D_t\langle J_s, J_s^{-}\rangle(1) \\ &= \langle D_tJ_s(1), J_s^{-}(1)\rangle + |J_s^{-}(1)|^2 \end{aligned} \tag{6.10}$$

$$\begin{aligned} \Rightarrow \langle D_tJ_s(1), J_s^{-}(1)\rangle &= 2\frac{\langle J_s(1), \dot{c}(s,1)\rangle^2}{|\dot{c}(s,1)|^2} - |J_s^{-}(1)|^2 \\ &= \frac{\langle J_s(1), \dot{c}(s,1)\rangle^2}{|\dot{c}(s,1)|^2} \end{aligned} \tag{6.11}$$

So
$$\begin{aligned}\langle D_tJ_s(1), J_s(1)\rangle &= \langle D_tJ_s(1), J_s^-(1)\rangle + \langle D_tJ_s(1), J_s^\perp(1)\rangle \\ &= \langle D_tJ_s(1), J_s^-(1)\rangle + \langle D_t(J_s^\perp)(1), J_s^\perp(1)\rangle \\ &= \frac{< J_s(1), \dot{c}(s,1) >^2}{|\dot{c}(s,1)|^2} + f(|\dot{c}(s,1)|)|J_s^\perp(1)|^2 \\ &= \frac{\langle J_s(1), \dot{c}(s,1)\rangle^2}{|\dot{c}(s,1)|^2} + f(|\dot{c}(s,1)|)|J_s(1)|^2 - f(|\dot{c}(s,1)|)\frac{\langle J_s(1), \dot{c}(s,1)\rangle^2}{|\dot{c}(s,1)|^2} \\ &= f(|\dot{c}(s,1)|)|J_s(1)|^2 + (1 - f(|\dot{c}(s,1)|))\frac{\langle J_s(1), \dot{c}(s,1)\rangle^2}{|\dot{c}(s,1)|^2} \\ &= f(d(\gamma(s), m))|\dot{\gamma}(s)|^2 + (1 - f(d(\gamma(s), m))\frac{\langle J_s(1), \dot{c}(s,1)\rangle^2}{|\dot{c}(s,1)|^2}\end{aligned} \tag{6.12}$$

From (6.6) and (6.12), we have

$$\begin{aligned}&\frac{d^2}{ds^2}F(\gamma(s)) = 2\int \langle D_tJ_s(1), J_s(1)\rangle Q(dm) \\ =&2\int \left(f(d(\gamma(s), m))|\dot{\gamma}(s)|^2 + (1 - f(d(\gamma(s), m))\frac{\langle J_s(1), \dot{c}(s,1)\rangle^2}{|\dot{c}(s,1)|^2}\right) Q(dm)\end{aligned} \tag{6.13}$$

Substituting $s = 0$ in (6.13), we get

$$\begin{aligned}\frac{d^2}{ds^2}F(\gamma(s))|_{s=0} &= 2\int \left(f(|y|)|\dot{\gamma}(0)|^2 + (1 - f(|y|))\frac{\langle \dot{\gamma}(0), y\rangle^2}{|y|^2}\right) Q^\phi(dy) \\ &= 2|\dot{\gamma}(0)|^2 Ef(|Y_1|) + 2E\left((1 - f(|Y_1|))\frac{\langle \dot{\gamma}(0), Y_1\rangle^2}{|Y_1|^2}\right)\end{aligned} \tag{6.14}$$

If we take $\dot{\gamma}(0) = \sum V^i\partial_i$, then we have

$$\sum V^iV^j\Lambda_{ij} = \frac{d^2}{ds^2}F(\gamma(s))|_{s=0} \tag{6.15}$$

So taking $\dot{\gamma}(0) = \partial_i$, we have

$$\begin{aligned}&|\dot{\gamma}(0)|^2 = 1 \\ &\Lambda_{ii} = 2Ef(|Y_1|) + 2E\left(\frac{1 - f(|Y_1|)}{|Y_1|^2}\right)(Y_1^i)^2\end{aligned} \tag{6.16}$$

Next taking $\dot{\gamma}(0) = \partial_i + \partial_j$, $i \neq j$ we get

$$
\begin{aligned}
&|\dot{\gamma}(0)|^2 = 2 \\
&\frac{d^2}{ds^2}F(\gamma(s))|_{s=0} = 4Ef(|Y_1|) + 2E\left((1 - f(|Y_1|))\frac{(Y_1^i + Y_1^j)^2}{|Y_1|^2}\right) \\
&= \Lambda_{ii} + 2\Lambda_{ij} + \Lambda_{jj} \\
&\Rightarrow \Lambda_{ij} = 2E\left(\frac{1 - f|Y_1|}{|Y_1|^2}\right)Y_1^iY_1^j
\end{aligned}
$$

This gives the expression for Λ and hence proves 4. □

To construct asymptotic confidence set for the population mean μ_I, we can consider the statistic

$$T_n = d_g^2(\mu_{nI}, \mu_I)$$

Taking $\phi = Exp_{\mu_I}^{-1}$ in Proposition 6.1, $T_n = \|\mu_n\|^2$.
Then from Proposition 6.1,

$$nT_n \xrightarrow{\mathcal{L}} \sum_{i=1}^{d} \lambda_i Z_i^2 \tag{6.17}$$

where $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_d$ are the eigen values of $\Lambda^{-1}\Sigma\Lambda^{-1}$ and $Z_1, \ldots, Z_d$ iid $N(0, 1)$.
So an asymptotic level $(1 - \alpha)$ confidence set for μ_I can be given by:

$$\{\mu_I : nd_g^2(\mu_{nI}, \mu_I) \leq \hat{c}_{1-\alpha}\} \tag{6.18}$$

where $\hat{c}_{1-\alpha}$ is the estimated upper $(1 - \alpha)$ quantile of the distribution of $\sum_{i=1}^{d} \hat{\lambda}_i Z_i^2$, where $\hat{\lambda}_i$ are the eigen values estimated from the sample $X_1, \ldots, X_n$ using Theorem 6.2 and $(Z_1, Z_2, \ldots)$ is a sample of iid $N(0, 1)$ independent of the sample $X_1, \ldots, X_n$.
The corresponding bootstrapped confidence region is given by

$$\{\mu_I : nd_g^2(\mu_{nI}, \mu_I) \leq c^*_{1-\alpha}\} \tag{6.19}$$

where $c^*_{(1-\alpha)}$ is the upper $(1 - \alpha)$-quantile of the bootstrapped values of the statistic nT_n.

7 Applications

Now we will apply the above results to some Riemannian manifolds including the Planer Shape Space.

7.1 Unit sphere S^{k-1}

At each $p \in S^{k-1}$, endow the tangent space $T_p = \{v \in \mathbb{R}^k : v.p = 0\}$ with the metric tensor $g_p : T_p \times T_p \to \mathbb{R}$ as the restriction of the scalar product at p of the tangent space of $\mathbb{R}^k$: $g_p(v_1, v_2) = \langle v_1, v_2 \rangle$. Then g is a smooth metric tensor on the tangent bundle $TS^{k-1} = \{(p, v) : p \in S^{k-1}, v \in \mathbb{R}^k : v.p = 0\}$. The geodesics are the big circles,

$$\gamma_{p,v}(t) = (\cos t)p + (\sin t)\frac{v}{\|v\|}, -\pi < t \leq \pi \tag{7.1}$$

The exponential map, $Exp : T_p \to S^{k-1}$ is

$$\begin{aligned} &Exp_p(0) = p, \\ &Exp_p(v) = \cos(\|v\|)p + \sin(\|v\|)\frac{v}{\|v\|} \quad (v \in T_p) \end{aligned} \tag{7.2}$$

The cutlocus of p is $C(p) = \{-p\}$. The inverse of the Exponential map on $S^{k-1} \setminus \{-p\}$ into T_p is

$$\begin{aligned} &Exp_p^{-1}(q) = \frac{arccos\langle p, q \rangle}{\sqrt{1 - \langle p, q \rangle^2}}[q - \langle p, q \rangle p] \quad (q \neq p, -p), \\ &Exp_p^{-1}(p) = 0 \end{aligned} \tag{7.3}$$

The geodesic distance d_g on S^{k-1} is

$$d_g(p, q) = |arccos\langle p, q \rangle| \in [0, \pi]$$

This space has constant sectional curvature 1. So if Q is a probability measure on S^{k-1}, Q has an intrinsic mean if its support is contained in a geodesic ball of radius at most $\pi/4$. (See Proposition 5.1). In this case the sample Intrinsic mean (i.e., any measurable selection from the sample intrinsic mean set) based on a random sample from Q is consistent.

In case Q has a mean μ_I, pick an orthonormal basis for $T_{\mu_I}S^d$: $\{v_1, \dots, v_d\}$. For $x \in S^{k-1}$, $|\langle x, \mu_I \rangle| < 1$, in Theorem 6.2, we have

$$\phi(x) = exp_{\mu_I}^{-1}(x) = \frac{arccos\langle x, \mu_I \rangle}{\sqrt{1 - \langle x, \mu_I \rangle^2}}[x - \langle x, \mu_I \rangle \mu_I] \tag{7.4}$$

Let $y = (y^1, \dots, y^{k-1}) = y^r v_r$ denote the normal coordinates of x. Then

$$\begin{aligned} y^r &= \frac{arccos\langle x, \mu_I \rangle}{\sqrt{1 - \langle x, \mu_I \rangle^2}}\langle x, v_r \rangle \quad r = 1, 2, \dots, k-1. \qquad (7.5) \\ |y| &= arccos\langle x, \mu_I \rangle = d_g(x, \mu_I) \end{aligned}$$

Q satisfies assumption (**A**), if all one dimensional curves have measure 0. This is true in particular if Q is absolutely continuous with respect to the volume measure. Then

$$f(p) = p\frac{cosp}{sinp} \tag{7.6}$$

$$\begin{aligned} \Lambda_{rs} = 2E[&\frac{1}{1 - \langle X_1, \mu_I \rangle^2}\left(1 - \frac{arccos\langle X_1, \mu_I \rangle}{\sqrt{1 - \langle X_1, \mu_I \rangle^2}}\langle X_1, \mu_I \rangle\right)\langle X_1, v_r \rangle \langle X_1, v_s \rangle \\ &+ \frac{arccos\langle X_1, \mu_I \rangle}{\sqrt{1 - \langle X_1, \mu_I \rangle^2}}\langle X_1, \mu_I \rangle \delta_{rs}] \quad 1 \leq r \leq s \leq k-1. \qquad (7.7) \end{aligned}$$

7.2 Planer Shape Space Σ_2^k

Consider first the complex projective space $\mathbb{C}P^d$: the space of all complex lines through the origin in $\mathbb{C}^{d+1}$. Consider the map

$$\begin{aligned} \pi &: \mathbb{C}S^d \to \mathbb{C}P^d \\ &z \mapsto \pi(z) = [z]; \\ &z \in \mathbb{C}^{d+1}, \ \|z\| = \sum_{j=1}^{d+1} |z_j|^2 = 1 \end{aligned}$$

This π is a Riemannian submersion. The tangent space $T_z\mathbb{C}S^d$ at **z** is the set of all vectors, **v** in $\mathbb{C}^{d+1}$ orthogonal to **z**. Here for $\mathbf{v}, \mathbf{w} \in \mathbb{C}^{d+1}$, $< \mathbf{v}, \mathbf{w} > = \mathrm{Re}(\mathbf{v}'\bar{\mathbf{w}})$.

Then for any $[z] \in \mathbb{C}P^d$, the tangent space $T_{[z]}\mathbb{C}P^d$ at $[z]$ is isomorphic with a subspace called the horizontal subspace of $T_z\mathbb{C}S^d$,

$$H_z = \{v \in \mathbb{C}^{d+1} \ : z^{'}\bar{v} = 0\}$$

Note that

$$T_z\mathbb{C}S^d = \{v \in \mathbb{C}^{d+1} \ : Re(z^{'}\bar{v}) = 0\}$$

It can be shown that $exp_{[z]} = \pi \circ exp_z \circ d\pi_z^{-1}$, and

$$d\pi_z^{-1}(exp_{[z]}^{-1}([w]) = \frac{r}{sinr}\{-zcosr + e^{i\theta}w\} \ \in H_z \tag{7.8}$$

where $z, w \in \mathbb{C}S^d$,

$$r = d_g([z],[w]) = \ arccos(|z^{'}\bar{w}|) \in \ [0, \frac{\pi}{2}) \tag{7.9}$$

and $e^{i\theta} = \dfrac{z^{'}\bar{w}}{|z^{'}\bar{w}|}$

Σ_2^k may be identified with the set of all complex lines in $H^{k-1} \equiv \{\mathbf{w} \in \mathbb{C}^k \setminus \{0\} : \sum_1^k w_j = 0\}$, with a k-ad $\mathbf{z} = (z_1, z_2, \ldots z_k) \equiv (x_1 + iy_1, \ldots x_k + iy_k)$ represented by $\mathbf{z}- < \mathbf{z} >$. So one may express the geodesics, geodesic distances, the exponential map and its inverse in Σ_2^k by simply taking $d = k-1$ above and replacing the k-ads $\mathbf{z}, \mathbf{w}$ by their preshapes (see Section 2.3). In view of the restriction to the complex hyperplane (of complex dimension k-1), the tangent space $T_{[z]}\Sigma_2^k$ at the shape $[z]$ is isomorphic with a subspace of H_z namely

$$\{v \in \mathbb{C}^k \ : z^{'}\bar{v} = 0, \ v^{'}\mathbf{1}_k = 0\}$$

Σ_2^k is then isomorphic (and isometric) to $\mathbb{C}P^{k-2}$.
This space has all sectional curvatures bounded between 1 and 4. The cut-locus of [p] is

$$\begin{aligned} C([p]) &= \{[x] : d_g([x],[p]) = \ arccos(|p^{'}\bar{x}|) = \frac{\pi}{2}\} \\ &= \{[x] : p^{'}\bar{x} = 0\} \end{aligned}$$

From Proposition 5.1, Q has an intrinsic mean, if its support is contained in a geodesic ball of radius at most $\frac{\pi}{8}$.
In case Q has intrinsic mean $[\mu]$, consider the coordinate patch $\phi = d\pi_\mu^{-1}(exp_{[\mu]}^{-1}[z])$

around $[\mu]$. From (7.8), we get

$$\begin{aligned}
&\phi : \Sigma_2^k \setminus C([\mu]) \to H_\mu \\
&\phi([z]) = \frac{r(z)}{sinr(z)}[-\mu cosr(z) + e^{i\theta(z)}z] \qquad (7.10)
\end{aligned}$$

where $z, \mu \in \mathbb{C}S^{k-1} \cap H^{k-1}$,

$$r(z) = d_g([z],[\mu]) = \ arccos(|\mu^{'}\bar{z}|) \in \ [0, \frac{\pi}{2})$$

and $e^{i\theta(z)} = \dfrac{\mu^{'}\bar{z}}{|\mu^{'}\bar{z}|}$.

Q satisfies assumption (**A**) if it is absolutely continuous wrt the volume measure on the shape space. If $\{[X_1], [X_2], \ldots, [X_j]\}$ is an iid sample from Q, X_j's being the pre-shapes, let $Y_j = \phi(X_j)$. Then in Theorem 6.2,

$$\begin{aligned}
Dh(Y_1, \mu) &= -2Y_1 \\
&= \frac{-2r_1}{sinr_1}[-cosr_1\mu + e^{i\theta_1}X_1] \qquad (7.11) \\
\text{where } r_1 &= arccos(|\mu^{'}\bar{X}_1|) \\
\text{and } e^{i\theta_1} &= \frac{\mu^{'}\bar{X}_1}{cosr_1}. \\
E(Dh(Y_1, \mu)) &= 0
\end{aligned}$$

So to find μ, we need to find the zeros of the function: $\mu \mapsto E(Dh(Y_1, \mu))$. This is equivalent to finding the fixed points of the map

$$\begin{aligned}
&f : \pi^{-1}\Sigma_2^k \to \mathbb{C}S^{k-1} \\
&\quad \mu \mapsto exp_\mu d\pi^{-1}_{\pi(\mu)}(-gradF(\pi(\mu)))
\end{aligned}$$

Here exp_μ is the exponential map on $\mathbb{C}S^{k-1}$,

$$\begin{aligned}
&exp_\mu : \ T_\mu\mathbb{C}S^{k-1} \to \mathbb{C}S^{k-1} \\
&\quad exp_\mu(v) = (cos|v|)\mu + \frac{sin|v|}{|v|}v. \\
&\quad \text{So } f(\mu) = (cos|v|)\mu + \frac{sin|v|}{|v|}v \qquad (7.9) \\
&\quad \text{where } v = 2E\frac{r_1}{sinr_1}(-cosr_1\mu + e^{i\theta_1}X_1)
\end{aligned}$$

A result (Theorem 4, LE [3]) says that if the support of Q is contained in a geodesic ball of radius at most $\frac{3\pi}{40}$, then f has a unique fixed point μ, and then $[\mu]$ is the intrinsic mean of Q.

The above result can be used to find the intrinsic sample mean by replacing Q by the empirical distribution, Q_n. That is, start with some $\theta_0 \in \mathbb{C}S^{k-1} \cap H^{k-1}$ and compute θ_m iteratively:

$$\begin{aligned} \theta_{m+1} &= f(\theta_m); \;\; m = 0, 1, 2, \dots \\ &= (cos|v_m|)\theta_m + \frac{sin|v_m|}{|v_m|} v_m \qquad (7.10) \\ \text{where } v_m &= 2\sum_{j=1}^{n} \frac{1}{n} \frac{r_j}{sinr_j}(-cosr_j\theta_m + e^{i\theta_j} X_j) \\ r_j &= arccos(|\theta_m^{'} \bar{X}_j|) \\ e^{i\theta_j} &= \frac{\theta_m^{'}\bar{X}_j}{cosr_j} \end{aligned}$$

If all the sample points are in a geodesic ball of radius at most $\frac{3\pi}{40}$, then the above algorithm converges to μ_n, $[\mu_n]$ being the sample intrinsic mean, whetever θ_0 we start with in that ball. We may take $[\theta_0]$ to be the sample extrinsic mean.

For the skull data, the female sample is contained in a geodesic ball of radius 0.0703 while the male sample is contained in a geodesic ball of radius 0.0855 around their respective sample extrinsic means. Since both radii are $<< 3\pi/40$, their sample intrinsic means exist and the above algorithm converges to that. The geodesic distance between the extrinsic and intrinsic means come to the order of 10^{-6}:

$$\begin{aligned} d_g(\mu_{nE}, \mu_{nI}) &= 5.5395e - 07 \\ d_g(\mu_{mE}, \mu_{mI}) &= 1.9609e - 06 \end{aligned}$$

Here μ_{nE} and μ_{mE} denote the preshapes of the extrinsic sample means of the females and males respectively while μ_{nI} and μ_{mI} denote the corresponding sample intrinsic means.

Using the sample estimates, we can construct 95% asymptotic confidence regions of the population intrinsic means as in (6.19). They are

$$\text{Females: } \{[\mu_1] : nd_g^2(\mu_{nI}, \mu_1) \leq 0.0003247\} \tag{7.10}$$

$$\text{Males: } \{\mu_2 : md_g^2(\mu_{mI}, \mu_2) \leq 0.0004691\} \tag{7.11}$$

Here $n = 30$ and $m = 29$ are the female and male sample sizes; μ_1 and μ_2 are the preshapes of their population intrinsic means.

The confidence regions in (7.10) and (7.11) can be used to to test if the male and female populations have same intrinsic mean shapes. We accept H_0: $[\mu_1] = [\mu_2]$ if the regions overlap that is if

$$d_g(\mu_{nI}, \mu_{mI}) < \sqrt{\frac{0.0003247}{n}} + \sqrt{\frac{0.0004691}{m}}$$

For this sample, $d_g(\mu_{nI}, \mu_{mI}) = 0.0587$ while $\sqrt{\frac{0.0003247}{n}} + \sqrt{\frac{0.0004691}{m}} = 0.0073$. So we reject H_0.

8 Conclusion

There are many outstanding statistical problems in shape analysis which remain unresolved. One of them is a proper analysis of 3-D shape spaces and of distributions on them. Another is the study of time series for evolution of the distribution of shapes, in discrete and in continuous time. For 2-D shape spaces and other Riemannian manifolds, a mathematical problem of some significance is to find broad conditions for the uniqueness of the intrinsic mean.

9 Acknowledgements

I am very thankful to my advisor, Dr. Rabi Bhattacharya for guiding me through out my research. I would also like to thank other professors for their thoughtful and constructive suggestions.

10 References

[1] BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large Sample Methods of Intrinsic and Extrinsic Sample Means on Manifolds-I. *Ann. Statist.*31 1-29.

[2] BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large Sample Methods of Intrinsic and Extrinsic Sample Means on Manifolds-II. *Ann. Statist.*33 1225-1259.

[3] KARCHAR, H. (1977). Riemannian Center of Mass & Mollifier Smoothing. *Comm. on Pure & Applied Math.*XXX 509-541.

[4] LE, HUILING (2001). Locating Frechet Means with Application to Shape Spaces. *Adv. Appl. Prob.*33 324-338.

[5] PENNEC, XAVIER (1999). Probabilities and Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *NSIP'99.*

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

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

[8] LEE, J.M. (1997). Riemannian Manifolds an introduction to Curvature. *Springer.*

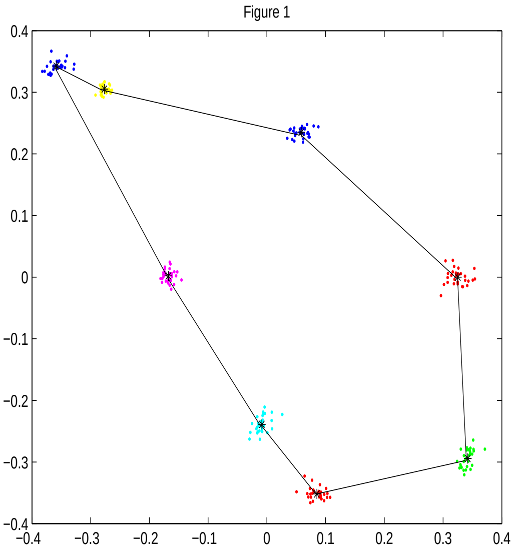

Figure 1: Procrustes Coordinates of Female landmarks.
* denotes coordinates of the mean shape.

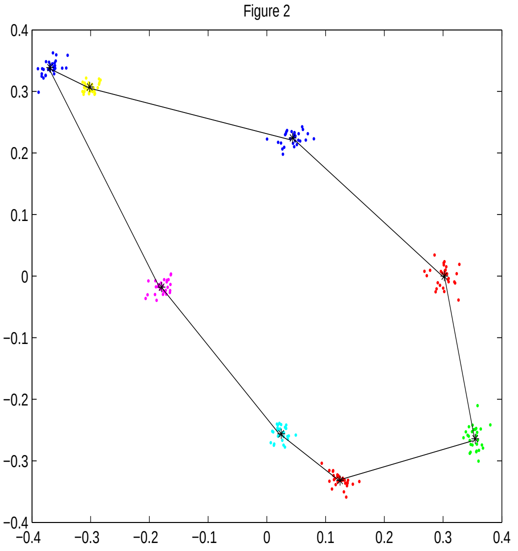

Figure 2: Procrustes Coordinates of Male landmarks.
* denotes coordinates of the mean shape.

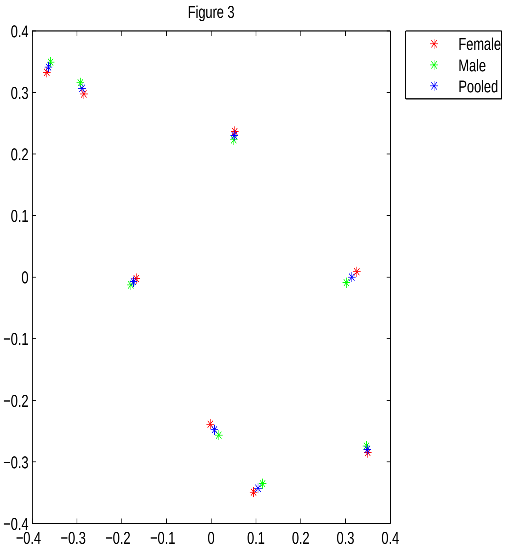

Figure 3: Procrustes Coordinates of sample means.