Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions[1]

By Harrie Hendriks and Zinoviy Landsman

University of Nijmegen and Haifa University

Abstract

In a previous investigation we studied some asymptotic properties of the empirical mean location on submanifolds of Euclidean space. The empirical mean location generalizes least squares statistics to smooth compact submanifolds of Euclidean space.

In this paper these properties are put into use. Tests for hypotheses about mean location are constructed and confidence regions for mean location are indicated. We study the asymptotic distribution of the testing statistic. The problem of comparing mean locations for two samples is analyzed.

Special attention was paid to observations on Stiefel manifolds including the orthogonal group $O(p)$ and spheres $\mathbf{S}^{k-1}$, and special orthogonal groups $SO(p)$. The results also are illustrated with our experience with simulations.

1 Introduction

Let X be a random variable with values in a C^∞-smooth m-dimensional compact manifold $\mathbf{M} \subset \mathbb{R}^k$. The concept of *mean location* as a point on the manifold minimizing mean square distance to X is the natural generalization of expectation vector in $\mathbb{R}^k$, more explicitly, mean location has been defined in Hendriks, Janssen & Ruymgaart (1992) (see also Hendriks (1991)) as

$$\mu = \underset{x \in \mathbf{M}}{\operatorname{argmin}} \; E(\|X - x\|^2). \tag{1}$$

Given n independent copies $X_1, \ldots, X_n$ of X, the empirical mean location is similarly defined as

$$\mu_n = \underset{x \in \mathbf{M}}{\operatorname{argmin}} \; \frac{1}{n} \sum_{i=1}^{n} \|X_i - x\|^2. \tag{2}$$

If $\mathbf{M} = \mathbf{S}^{k-1}$, the unit sphere in $\mathbb{R}^k$, the minimization in (1) and (2) can easily be carried out. Let $E(X)$ be the ordinary population mean or expectation vector; then we have

$$\mu = \frac{E(X)}{\|E(X)\|}, \text{ provided that } E(X) \neq 0.$$

[1] AMS 1991 *subject classifications.* Primary 62H11, 62H15; secondary 62G15, 62G20.

Key words and phrases. Mean location, degenerate normal distribution, sample Weingarten mapping, cut-locus, Stiefel manifold, test, confidence interval, asymptotic, acceptance-rejection method, simulation.

Similarly, let $\overline{X}_n$ be the ordinary sample mean vector; then we have

$$\mu_n = \frac{\overline{X}_n}{\|\overline{X}_n\|}, \text{ provided that } \overline{X}_n \neq 0,$$

so μ and μ_n correspond to mean direction and sample mean direction, respectively. Mean direction has been studied in Mardia (1975a), Watson (1983a), Fisher, Lewis, Embleton (1987) and others. In Hendriks, Landsman & Ruymgaart(1996) the asymptotic behavior of sample mean direction was considered without any restrictions on rotational symmetry for the underlying distribution of X. Under the assumptions, that $E(X) \neq 0$ and the covariance matrix $\mathrm{Var}(X)$ of X is positive definite, it was shown that the $n^{1/2}$ - normalized difference between sample mean location and mean location, $n^{1/2}(\mu_n - \mu)$, has a degenerate normal limiting distribution $\mathcal{N}(0, \widehat{\Sigma})$ with support on the tangent space to the sphere at the point μ, where $\widehat{\Sigma}$ is a degenerate covariance matrix of rank $k-1$. Further investigation unveiled that its normal component N_n has order $O(n^{-1})$ and that $-2nN_n$ has a χ^2-type limiting distribution which is the sum of squares of $(k-1)$ independent normal variables with zero mean and possibly different variances.

In Hendriks and Landsman (1996a) it was shown that, mutatis mutandis, these results remain true for arbitrary smooth submanifolds of Euclidean spaces. It requires the introduction of some differential geometric notions, namely *cut-locus* and *Weingarten mapping.* If a random variable X has values in the sphere $\mathbf{S}^{k-1}$, mean location μ is uniquely defined for any $a = E(X)$, except $a = 0$. For other manifolds under consideration, there exists a (possibly non finite) set of such exception points, named *cut-locus* by Thom (1972). The exact definitions and some more details will be given in Section 2. Under the assumptions that $E(X)$ is not in the cut-locus, and $\mathrm{Var}(X)$ is positive definite, it was shown that for samples with mean value $\overline{X}_n$, lying outside the cut-locus, μ_n is uniquely defined and $n^{1/2}(\mu_n - \mu)$ is asymptotically degenerate normally distributed $\mathcal{N}(0, \widehat{\Sigma})$; the degeneracy is such that the limiting distribution is supported by the tangent space $T_\mu\mathbf{M}$ to the manifold $\mathbf{M}$ at the point μ. The degenerate covariance matrix $\widehat{\Sigma}$ now has rank m, being the dimension of $\mathbf{M}$ and may be obtained in terms of the population covariance matrix Σ and the Weingarten mapping. The projection of the vector $(\mu_n - \mu)$ to any unit normal vector v_μ, ${v_\mu}^T(\mu_n - \mu)$, was shown to be of order $O(n^{-1})$, and $2n{v_\mu}^T(\mu_n - \mu)$ is asymptotically distributed as the linear combination of m independent squared Gaussian $\mathcal{N}(0,1)$ variables.

These results enable the construction in Section 3 of nonparametric asymptotic tests for the hypothesis H_0 that the population mean location equals μ and of confidence regions for μ. In their construction, we use the addition of a desingularizing term to $\widehat{\Sigma}$ as announced in Hendriks and Landsman (1996b). This term is supported by the normal space $N_\mu\mathbf{M}$ of vectors orthogonal to the tangent space $T_\mu\mathbf{M}$, and together with $\widehat{\Sigma}$ leads to a nonsingular form. Its influence is asymptotically negligeable since the orthogonal component of $\mu_n - \mu$ is of order $O_L(n^{-1})$. It not only allows avoiding the somewhat inconvenient Moore inversion, but enables to test consistently against μ-alternatives that differ by a normal vector and leads to confidence regions asymptotically retracting to a single point. In two-sample testing this modification plays an even more essential role in circumventing problems due to discontinuity of Moore inversion.

Section 4 is devoted to illustrate our result for some special manifolds: Stiefel manifolds

$\mathbf{V}_{p,r}$, $r \leq p$, and special orthogonal groups $SO(p)$. Stiefel manifold, being the manifold of orthonormal r-frames in p-dimensions (Downs (1972)) is most important. Special attention was paid to the cases of spheres $\mathbf{S}^{k-1} = \mathbf{V}_{k,1}$, and orthogonal groups $O(p) = \mathbf{V}_{p,p}$.

Section 5 is devoted to the two sample problem, that is, to test for the significance of the difference between the sample mean locations of two samples from possibly different probability distributions. In Section 6 we report on our experience in simulations. They show that the theoretical asymptotic results and the simulated results are very close, already for moderate sample sizes. Much like Student distribution converges to normal distribution for increasing number of degrees of freedom. Of course, we could have worked out the example vectorcardiogram data used in Downs (1972) and Prentice (1986). We chose to illustrate by simulation, not only for lack of availability of the vectorcardiogram data, but especially for the superiority of simulated data as the exact distribution is supposedly known.

Examples of observations on manifolds are arising in many problems (see Mardia(1975a), Jupp and Mardia (1979). Mardia (1988)). In vectorcardiography and astronomy one comes across observations on Stiefel manifolds $\mathbf{V}_{3,2}$. The orientation of so-called vectorcardiogram QRS loop (Downs (1972), Mardia and Khatri (1977), Khatri and Mardia (1977), Prentice (1986), Prentice (1989)) and the orientation of a comet's orbit (Mardia (1975b), Jupp and Mardia (1979)) can be specified by elements of $\mathbf{V}_{3,2}$. Some new interesting distributional results on Stiefel manifolds were presented by Chikuse (1990a).

2 Asymptotics of the sample mean location

Recall, the *cut-locus* of $\mathbf{M}$ is the set $\mathbf{C}$ of points $y \in \mathbb{R}^k$ for which the function L_y : $\mathbf{M} \to \mathbb{R} : x \mapsto \|x - y\|^2$ has not a unique minimum or attains its minimum in a point x, which is a degenerate critical point of the function L_y. A point x is a critical point of the function f on the submanifold $\mathbf{M}$ if in local coordinates $u^1, ..., u^m$, $\frac{\partial f}{\partial u^j}|_x = 0$ and is a degenerate critical point if, moreover, the matrix of second derivatives $\left(\frac{\partial^2 f}{\partial u^i \partial u^j}|_x\right)_{i,j=1}^m$ is a singular matrix. Denote by $\mathbf{U} = \mathbb{R}^k \backslash \mathbf{C}$. Then from Hendriks (1990, 1992), $\mathbf{C}$ is nowhere dense and $\mu_k(\mathbf{C}) = 0$, where μ_k denotes Lebesgue measure in $\mathbb{R}^k$, and the nearest point mapping $\Phi : \mathbf{U} \to \mathbf{M}$, defined as

$$\Phi(y) = \underset{x \in \mathbf{M}}{\operatorname{argmin}}\ L_y(x),$$

is C^∞ - differentiable on $\mathbf{U}$. For definiteness, we arbitrarily extend Φ to a measurable mapping defined on $\mathbb{R}^k$, such that (if possible) L_y assumes its minimum value at $\Phi(y)$.

Let $a = EX, \overline{X}_n = n^{-1} \sum_{i=1}^n X_i$ be the ordinary mean and sample mean vectors. Mean location and sample mean location are as follows

$$\mu = \Phi(a), \text{ if } a \in \mathbf{U}; \quad \mu_n = \Phi(\overline{X}_n), \text{ if } \overline{X}_n \in \mathbf{U}.$$

In the sequel it is supposed that $a \in \mathbf{U}$. Let $\mu = \Phi(a)$. Denote by $T_\mu \mathbf{M}$ the tangent space to the manifold $\mathbf{M}$ at the point $\mu \in \mathbf{M}$. In the usual way we will consider $T_\mu \mathbf{M}$ as a linear subspace of $T_\mu \mathbb{R}^k$ and we will identify $T_\mu \mathbb{R}^k$ with $\mathbb{R}^k$, using the structure of $\mathbb{R}^k$ as a vector space. Let $N_\mu \mathbf{M}$ be the vector space of vectors orthogonal to $T_\mu \mathbf{M}$. Notice that $a - \mu \in N_\mu \mathbf{M}$.

We will need to study the Jacobian of the function Φ at the point a, $\Phi'_a : T_a\mathbb{R}^k \to T_\mu\mathbf{M}$. Recall that any normal vector $n_\mu \in N_\mu\mathbf{M}$ determines a linear map, the *Weingarten mapping* (Kobayashi & Nomizu (1969), p. 15) given by

$$A_{n_\mu} : T_\mu\mathbf{M} \to T_\mu\mathbf{M} : A_{n_\mu}(w_\mu) = -\tan_\mu(D_{w_\mu}(n)),$$

where $n : \mathbf{M} \to \mathbb{R}^k$ is any smooth mapping such that $n(\alpha) \in N_\alpha\mathbf{M}$, for all $\alpha \in \mathbf{M}$, and such that $n(\mu) = n_\mu$. $D_{w_\mu}(\cdot)$ denotes coordinate wise differentiation with respect to the direction $w_\mu \in T_\mu\mathbf{M}$, and $\tan_\mu(\cdot)$ denotes the orthogonal projection onto $T_\mu\mathbf{M}$. Both $\tan_\mu$ and the Weingarten mapping A_{n_μ} are self adjoint (symmetric) with respect to the Euclidean inner product. Let Id_μ stand for the identity mapping of $T_\mu\mathbf{M}$ and E_k stand for the identity mapping of $\mathbb{R}^k$. The next lemma explains the relevance of the Weingarten mapping. It appears as a partial result, Equation (6), in the proof of Theorem 1 of Hendriks & Landsman (1996a).

Lemma 1 *Let $a \in \mathbf{U}$, and $\mu = \Phi(a)$. Then $\mathrm{Id}_\mu - A_{a-\mu} : T_\mu\mathbf{M} \to T_\mu\mathbf{M}$ is non singular. Let $B_\mu = (\mathrm{Id}_\mu - A_{a-\mu})^{-1}$, then*

$$\Phi'_a = B_\mu \tan_\mu = (\mathrm{Id}_\mu - A_{a-\mu})^{-1}\tan_\mu.$$

The following theorems describe the asymptotic behavior of $\sqrt{n}(\mu_n - \mu)$ and its projection on a normal vector v_μ and were proved in Hendriks and Landsman (1996a). Let Σ denote the covariance matrix of X, and let

$$\hat{\Sigma} = (B_\mu \tan_\mu)\, \Sigma\, (B_\mu \tan_\mu)^T$$

Theorem 1 *Suppose $EX \in \mathbf{U}$ and Σ positive definite. For $n \to \infty$, $\sqrt{n}(\mu_n - \mu)$ converges in distribution to the degenerate Gaussian distribution $\mathcal{N}(0, \hat{\Sigma})$, with covariance matrix $\hat{\Sigma}$ of rank m. The support of this distribution equals the tangent space $T_\mu\mathbf{M}$.*

Theorem 2 *Suppose $EX \in \mathbf{U}$ and Σ positive definite. Then the random variable $2nv_\mu^T(\mu_n - \mu)$ asymptotically has the same distribution as a linear combination of the squares of m independent standard normally distributed random variables $\xi_1, \ldots, \xi_m$. The sign of each coefficient of ξ_j^2, $j = 1, ..., m$, corresponds to the sign of a nonzero eigenvalue of the matrix A_{v_μ}.*

3 Asymptotics of the mean location test statistic and confidence regions

Given a sample $X_1, X_2, \ldots, X_n$ of iid random variables one may consider the sample Weingarten mapping $A_{\overline{X}_n - \mu_n}$ associated with the normal vector $\overline{X}_n - \mu_n \in N_{\mu_n}\mathbf{M}$ and the related mapping $G_n = (Id_{\mu_n} - A_{\overline{X}_n - \mu_n})\tan_{\mu_n} + (E_k - \tan_{\mu_n}) = E_k - A_{\overline{X}_n - \mu_n}\tan_{\mu_n}$. The linear map $G_n : \mathbb{R}^k \to \mathbb{R}^k$ is non singular if $\overline{X}_n \in \mathbf{U}$, and it is self adjoint.

Lemma 2 *Suppose* $a = E(X) \in \mathbf{U}$. *Then for* $n \to \infty$, $G_n \xrightarrow{P} G_a = (Id_\mu - A_{a-\mu})\tan_\mu + (E_k - \tan_\mu)$ *as a sequence in the vector space of* $k \times k$*-matrices which can be identified with* $\mathbb{R}^{k^2}$.

Lemma 2 follows from well-known theorems for convergence in probability (see, for example, Fuller (1976), Ch. 5.1). In particular $\overline{X}_n \xrightarrow{P} a$, and since Φ is continuous at a, $\mu_n \xrightarrow{P} \mu$. Moreover the Weingarten mapping A_{n_μ} depends continuously on the parameter n_μ (and μ).

Let $V_n = n^{-1} \sum_{i=1}^n (X_i - \overline{X}_n)(X_i - \overline{X}_n)^T$ be the sample covariance matrix, and let $\tilde{V}_n = \tan_{\mu_n} V_n \, (\tan_{\mu_n})^T + (E_k - \tan_{\mu_n})$. Consider the test statistics

$$M_n = \sqrt{n} G_n(\mu_n - \mu), \quad T_n = \tilde{V}_n^{-1/2} M_n.$$

Here $\tilde{V}_n^{-1/2}$ is understood to be the symmetric positive definite square root of the inverse of $\tilde{V}_n$. More generally, if $\tilde{V}_n$ happens to be singular, the positive semi-definite square root of the Moore inverse of $\tilde{V}_n$ could be chosen. Let Σ denote the covariance matrix of X, which is supposed to be positive definite. With probability approaching 1 for large n, V_n will be non singular, as $V_n \xrightarrow{P} \Sigma$. The following theorem provides a test for the hypothesis H_0 that the population mean location equals μ.

Theorem 3 *Suppose* $EX \in \mathbf{U}$ *and* Σ *non singular. Suppose the hypothesis* H_0 *is satisfied. Then, for* $n \to \infty$,

1. $M_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\mu)$, *where* $\mathcal{N}(0, \Sigma_\mu)$ *is the degenerate Gaussian distribution with covariance matrix* $\Sigma_\mu = \tan_\mu \Sigma (\tan_\mu)^T$ *of rank* m. *The support of* $\mathcal{N}(0, \Sigma_\mu)$ *equals the tangent space* $T_\mu \mathbf{M}$.

2. $T_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tan_\mu)$, *where* $\mathcal{N}(0, \tan_\mu)$ *is the degenerate normal distribution with support on* $T_\mu \mathbf{M}$ *and covariance matrix equal to the Euclidean inner product on* $T_\mu \mathbf{M}$.

3. $t_n = T_n^T T_n = M_n^T \tilde{V}_n^{-1} M_n \xrightarrow{\mathcal{D}} \chi^2_m$.

Notice that Statement 1 provides a test for H_0 with known limit size, if EX is unknown, but known to lie in $\mathbf{U}$, and Σ known. Statement 2, as well as 3, provides a test for H_0 with known limit size, if EX is unknown, but known to lie in $\mathbf{U}$, and Σ unknown, but known to be non singular.

Proof of Theorem 3. According to Theorem 1, $\sqrt{n}(\mu_n - \mu)$ converges in distribution to $\mathcal{N}(0, \widehat{\Sigma})$. Since G_n converges in probability to G_a and $G_a B_\mu = \mathrm{Id}_\mu$, it follows that $M_n = G_n(\sqrt{n}(\mu_n - \mu))$ converges in probability to $\mathcal{N}(0, G_a \widehat{\Sigma} G_a^T)$ with covariance matrix $G_a \widehat{\Sigma} G_a^T = \tan_\mu \Sigma \tan_\mu = \Sigma_\mu$. Moreover, $V_n \xrightarrow{P} \Sigma$ and $\mu_n \xrightarrow{P} \mu$. Therefore $\tilde{V}_n \xrightarrow{P} \Sigma_\mu + (E_k - \tan_\mu)$, and $T_n = \tilde{V}_n^{-1/2} M_n$ converges in distribution to the degenerate normal distribution with covariance matrix $\tan_\mu$. The third claim follows immediately, since $\tan_\mu$ is an orthogonal projection matrix of rank m. □

Statement 3 of the Theorem can be used in the construction of confidence regions. Notice that $M_n^T \tilde{V}_n^{-1} M_n = n(\mu_n - \mu)^T G_n^T \tilde{V}_n^{-1} G_n(\mu_n - \mu)$, so that we obtain the following

Corollary 1 *Let c_α be the $(1-\alpha)$-quantile of the χ^2_m-distribution. Consider the region S_n of points μ satisfying the inequality*

$$(\mu_n-\mu)^T G_n^T \tilde{V}_n^{-1} G_n(\mu_n-\mu) \le c_\alpha/n$$

then asymptotically, S_n is a $(1-\alpha)$-confidence region. More precisely, if μ is the mean location, then

$$\mathbb{P}\{(\mu_n-\mu)^T G_n^T \tilde{V}_n^{-1} G_n(\mu_n-\mu) \le c_\alpha/n\} \to 1-\alpha, \ \textit{as } n\to\infty.$$

Remark 1 Notice the desingularization artifice – the addition of the $(E-\tan_{\mu_n})$ term – for the singular matrices involved in the formulation of the test statistics, leading to the entities $\tilde{V}_n$ and G_n. The corresponding quadratic form $G_n^T\tilde{V}_n^{-1}G_n$ is asymptotically positive definite, so that almost surely asymptotically the confidence region retracts to a single point, which we consider as a desirable property. Moreover, this term simplifies the proof of Statement 2 of Theorem 3, as Moore inversion is not a continuous operation so that the Continuous Mapping Theorem cannot be used directly. Also Statement 3 gives rise to an asymptotically consistent test of H_0. Otherwise the test would fail for μ-alternatives whose difference vector to the hypothetical μ are orthogonal to $T_\mu\mathbf{M}$. Notice, that according to Theorem 2 the normal part of $\mu_n-\mu$ is of order $O(\frac{1}{n})$, so that one could even replace the $(E-\tan_{\mu_n})$ term in $\tilde{V}_n$ by some scalar multiple $c_n(E-\tan_{\mu_n})$ with $c_n=o(n^{-1})$, for example $c_n=n^{-1/2}$, which would increase even more the power of the test at these alternatives.

4 Applications

In this section we will exhibit some important cases for which mainly the geometric ingredients are made explicit.

4.1 Stiefel Manifolds

Let $\mathrm{Mat}_{p,r}$ be the vector space of real $p\times r$, $r\le p$ matrices, with the inner product $(A,B)=\mathrm{Tr}\,A^TB$, so that $\|A-B\|^2=\mathrm{Tr}\,(A-B)^T(A-B)$. This corresponds to the Euclidean inner product with respect to the identification $\mathrm{Mat}_{p,r}=\mathbb{R}^{pr}$ (thus $k=pr$), by putting the matrix coefficients in one column of size pr. Let Sym_r be the vector space of symmetric $r\times r$ matrices, and Sym_r^+ the subset of semi positive definite symmetric $r\times r$ matrices. Let E_r denote the $r\times r$ unit matrix. Then $\mathbf{M}=\mathbf{V}_{p,r}\subset\mathrm{Mat}_{p,r}$ denotes the Stiefel manifold $\mathbf{V}_{p,r}=\{V\in\mathrm{Mat}_{p,r}|V^TV=E_r\in\mathrm{Sym}_r\}$ whose dimension is $m=pr-r(r+1)/2$. $\mathbf{M}$ can be considered as a compact submanifold of the vector space $\mathrm{Mat}_{p,r}$ given by the equation

$$f(X)=X^TX=E_r,$$

where f is considered as a mapping between vector spaces with inner product $f:\mathrm{Mat}_{p,r}\to\mathrm{Sym}_r$. Therefore a point in the Stiefel manifold is a $p\times r$ matrix with orthogonal columns of Euclidean norm 1. As a particular case $\mathbf{V}_{k,1}$ corresponds to the sphere $\mathbf{S}^{k-1}$, and $V_{p,p}$ equals the orthogonal group $\mathbf{O}(p)$.

Recall that the Stiefel manifold $\mathbf{M} = \mathbf{V}_{p,r}$ is a homogeneous space with respect to the Lie group $\mathbf{O}(p)$ where the action is defined as the restriction of the action by matrix multiplication

$$\mathbf{O}(p) \times \mathrm{Mat}_{p,r} \ni (g, X) \mapsto gX \in \mathrm{Mat}_{p,r}.$$

As a matter of fact $\mathbf{M}$ admits the larger symmetry group $\mathbf{O}(p) \times \mathbf{O}(r)$, where the action is defined by restriction of the action

$$(\mathbf{O}(p) \times \mathbf{O}(r)) \times \mathrm{Mat}_{p,r} \ni (g, h, X) \mapsto gXh^T \in \mathrm{Mat}_{p,r}.$$

It is clear that these symmetries are in fact isometries, namely $\|g(x-y)h^T\|^2 = \mathrm{Tr}\,(g(x-y)h^T)^T(g(x-y)h^T) = \mathrm{Tr}\,h(x-y)^T g^T g(x-y)h^T = \mathrm{Tr}\,h(x-y)^T(x-y)h^T = \mathrm{Tr}\,h^T h(x-y)^T(x-y) = \mathrm{Tr}\,(x-y)^T(x-y) = \|x-y\|^2$. Let $U \in \mathrm{Mat}_{p,r}$, then the elements X of $\mathbf{M}$ for which $L_U(X) = \|U - X\|^2 = (U - X, U - X)$ is minimal are characterized by the condition that $U = Xc$ where $c \in \mathrm{Sym}_r^+$ is the (semi) positive definite symmetric square root of $U^T U$. X is uniquely defined if c is non singular, thus if U is non singular. And in that case X is a non degenerate critical point of the function $L_U : \mathbf{M} \to \mathbb{R}$. If U is of rank less than r, unicity of X fails as there is a 1-1 correspondence between isometric (injective) mappings $Y : \mathrm{Ker}\,U \to (\mathrm{Im}\,U)^\perp$ and solutions V to the equation $U = Vc$, given by $Y = V|_{kerU}$. So the *cut-locus* $\mathbf{C}$ consists exactly of the $p \times r$-matrices U which are not of maximal rank, r. The nearest point mapping $\Phi : \mathrm{Mat}_{p,r} \backslash \mathbf{C} \to \mathbf{M}$ will be given by

$$\Phi(U) = U(U^T U)^{-1/2}, \text{ for } U \notin \mathbf{C}, \tag{3}$$

where $(U^T U)^{-1/2}$ is the positive definite symmetric square root of $U^T U$.
Notice that the linear mapping $Df|_X : \mathrm{Mat}_{p,r} \to \mathrm{Sym}_r$ is given by $Df|_X(H) = H^T X + X^T H$ and that its transpose (or rather adjoint with respect to the inner products) $(Df|_X)^T : \mathrm{Sym}_r \to \mathrm{Mat}_{p,r}$ is given by $(Df|_X)^T(c) = 2Xc$. So, for $\mu \in \mathbf{M}$, we have $Df|_\mu (Df|_\mu)^T(c) = 4c$. The orthogonal projection on the tangent space $T_\mu(\mathbf{M})$ will be given by

$$\tan_\mu(U) = U - (Df|_\mu)^T \left(Df|_\mu (Df|_\mu)^T\right)^{-1} Df|_\mu(U) = U - \frac{1}{2}\mu\left[\mu^T U + U^T \mu\right].$$

For $\mu \in \mathbf{M}$ a normal vector $n_\mu \in N_\mu(\mathbf{M}) \subset \mathrm{Mat}_{p,r}$ must lie in the range of $(Df|_\mu)^T$, so that $n_\mu = (Df|_\mu)^T(s) = 2\mu s$ for some $s \in \mathrm{Sym}_r$. Thus $2s = \mu^T n_\mu$ and therefore $\mu^T n_\mu$ is symmetric and $n_\mu = \mu[\mu^T n_\mu]$. Since $T_\mu(\mathbf{M})$ is precisely the orthogonal complement, it consists of the matrices $W \in \mathrm{Mat}_{p,r}$ for which $\mu^T W$ is anti-symmetric. To calculate the Weingarten mapping at the point μ in the normal direction n_μ, notice that it is extended to a field of normal vectors on $\mathbf{M}$ by $n_\alpha = \alpha[\mu^T n_\mu]$, $\alpha \in \mathbf{M}$. And the Weingarten mapping is defined as the mapping $A_{n_\mu} : T_\mu(\mathbf{M}) \to T_\mu(\mathbf{M})$ with $A_{n_\mu}(W) = -\tan_\mu(D_W(n)) = -\tan_\mu(W\mu^T n_\mu)$. Therefore for $W \in T_\mu(\mathbf{M})$,

$$\begin{aligned} A_{n_\mu}(W) &= -W\mu^T n_\mu + \frac{1}{2}\mu\left[\mu^T W \mu^T n_\mu + n_\mu^T \mu W^T \mu\right] \\ &= -\frac{1}{2}[E_p - \mu\mu^T] W n_\mu^T \mu - \frac{1}{2}[W n_\mu^T - n_\mu W^T]\mu \\ &= -W\mu^T n_\mu - \frac{1}{2}\mu W^T n_\mu + \frac{1}{2} n_\mu W^T \mu. \end{aligned}$$

Remark 2 Let $\mathbf{M} = O(p)$ be the set of orthogonal matrices. It coincides with the Stiefel manifold $V_{p,p}$. Then $\mu\mu^T = E_p$ and A_{n_μ} may be slightly simplified (cf. Hendriks and Landsman (1996a), Example ii)

$$A_{n_\mu}(W) = -\frac{1}{2}[Wn_\mu^T - n_\mu W^T]\mu = -\frac{1}{2}W\mu^T n_\mu - \frac{1}{2}n_\mu\mu^T W, \qquad W \in T_\mu(\mathbf{M}).$$

Remark 3 Let $\mathbf{M} = \mathbf{S}^{k-1}$ be the unit sphere in $\mathbb{R}^k$. It is a special case of Stiefel manifold $V_{k,1}$. In this case $\mathbf{C} = \{0\}$, $\mathbf{U} = \mathbb{R}^k \setminus \{0\}$ and $\Phi(a) = a/\|a\|$ for $a \in \mathbf{U}$. Let $a \in \mathbf{U}$ be the population mean then $\mu = a/\|a\|$, $\mu_n = \overline{X}_n/\|\overline{X}_n\|$, $\tan_{\mu_n} = E_k - \mu_n{\mu_n}^T$, and $G_n = \|\overline{X}_n\|\tan_{\mu_n} + \mu_n{\mu_n}^T$. See Hendriks and Landsman (1996a, 1996b).

4.2 Special orthogonal group $\mathbf{M} = SO(p)$

The discussion differs from the one on Stiefel manifolds only in the determination of the cut-locus and a necessary modification to the nearest point mapping Φ. Let $U \in \mathrm{Mat}_{p,p}$, then the elements X of $\mathbf{M}$ for which $L_U(X) = \|U - X\|^2 = (U - X, U - X)$ is stationary are characterized by the condition that $U = Xc$ where $c \in \mathrm{Sym}_p$ is a symmetric square root of $U^T U$. In order to have minimal distance, we need the square root with largest trace. This means that if $\det(U) > 0$, we can use the positive definite root. If $\det(U) < 0$, we must allow a negative root of the smallest eigenvalue of $U^T U$, which will give rise to non uniqueness if the smallest eigenvalue has multiplicity greater than 1. In the same vein, if $\det(U) = 0$ and the eigenvalue 0 has multiplicity greater than 1 there is not unicity of X. But if 0 is a simple eigenvalue of $U^T U$ the condition that $\det(X) = 1$ determines a single solution. Moreover one can prove that if $\det(U) > 0$ or both $\det(U) \leq 0$ and the lowest eigenvalue of $U^T U$ is simple, the minimum X is non degenerate. Therefore the cut-locus $\mathbf{C}$ is the set of $p \times p$ matrices U such that $\det(U) \leq 0$ and the lowest eigenvalue of $U^T U$ is not simple.

5 Two sample problem

In the problem of testing for the significance of the difference of the sample mean locations for two samples of observations on manifolds we have to face an important complication in comparison with observations on $\mathbb{R}^k$: the absence of translation invariance. This becomes apparent in the ambiguity of the choice of the supposed common mean location, occurring in the limit distributions of plausible test statistics. Moreover, we are confronted with the fact that the usual treatment in the literature is not so conclusive, even for observations in $\mathbb{R}$. Namely, the weak convergence claim

$$(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^{-1/2}(\overline{X}_1 - \overline{X}_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad \min(n_1, n_2) \to \infty,$$

where $\overline{X}_i, s_i^2$ are sample mean and variance of the sample of size n_i, $i = 1, 2$. Actually a usual proof refers to the well-known Slutsky theorem (see, for example, Bickel and Doksum (1977), §6.4.B, 6.4.C, A.14.9). This is the reason for us to give a more detailed treatment.

Let

$$X_{i,1}, ..., X_{i,n_i}, \quad i = 1, 2,$$

be independent samples on $\mathbf{M}$ with expectation points $a_i = EX_{i,1}$, covariance matrices Σ_i and mean locations μ_i, and suppose $a_i \in \mathbf{U}$ and Σ_i is non singular, $i = 1, 2$.

Denote by $G_i = G_{a_i}$, $B_i = B_{i,\mu_i} = (\mathrm{Id}_{\mu_i} - A_{a_i - \mu_i})^{-1}$ and $\sigma_i^2 = B_i \tan_{\mu_i} \Sigma_i \tan_{\mu_i} B_i^T + (E - \tan_{\mu_i})$, $\tilde{V}_i = \tan_{\mu_i} \Sigma_i \tan_{\mu_i} + (E - \tan_{\mu_i})$. In particular $\tilde{V}_i = G_i \sigma_i^2 G_i$. Let us introduce the sample statistics, for $i = 1, 2$: $\overline{X}_i = \overline{X}_{i,n_i} = n_i^{-1} \sum_{j=1}^{n_i} X_{i,j}$ – sample means, μ_{i,n_i} – sample mean locations, $V_{i,n_i} = n_i^{-1} \sum_{j=1}^{n_i} (X_{i,j} - \overline{X}_{n_i})(X_{i,j} - \overline{X}_{n_i})^T$ – sample variances. Set $G_{i,n_i} = (\mathrm{Id}_{\mu_{i,n_i}} - A_{\overline{X}_{n_i} - \mu_{i,n_i}}) \tan_{\mu_{i,n_i}} + (E_k - \tan_{\mu_{i,n_i}})$, $B_{i,n_i} = B_{i,\mu_{i,n_i}} = (\mathrm{Id}_{\mu_{i,n_i}} - A_{\overline{X}_{i,n_i} - \mu_{i,n_i}})^{-1}$, $\tilde{V}_{i,n_i} = \tan_{\mu_{i,n_i}} V_{i,n_i} \tan^T_{\mu_{i,n_i}} + (E_k - \tan_{\mu_{i,n_i}})$, $S_i^2 = B_{i,n_i} \tan_{\mu_{i,n_i}} V_{i,n_i} \tan^T_{\mu_{i,n_i}} B^T_{i,n_i} + (E_k - \tan_{\mu_{i,n_i}})$. In particular $\tilde{V}_{i,n_i} = G_{i,n_i} S_i^2 G_{i,n_i}$. Notice that all the mappings are symmetric. Moreover the mappings G_i and B_i are invertible and the matrices of σ_i^2 and $\tilde{V}_i$ are positive definite. In Theorem 4 we will see that, with probability approaching 1 for large n, these properties also will hold for G_{i,n_i}, B_{i,n_i} and S_i^2 and $\tilde{V}_{i,n_i}$.
The following theorem provides a test for the hypothesis $H_0 : \mu_1 = \mu_2$.

Theorem 4 *Suppose $a_i \in \mathbf{U}$, and Σ_i positive definite, $i = 1, 2$, and $\mu_1 = \mu_2$. Then*

1. $S_i^2 \xrightarrow{P} \sigma_i^2$, $G_{i,n_i} \xrightarrow{P} G_i$, $\tilde{V}_{i,n_i} \xrightarrow{P} \tilde{V}_i$, *as* $n_i \to \infty$,
2. $\xi_{i,n_i} = \sqrt{n_i} \tilde{V}_{i,n_i}^{-1/2} G_{i,n_i} (\mu_{i,n_i} - \mu_i) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tan_{\mu_i})$, *as* $n_i \to \infty$,
3. $T_{n_1,n_2} = (\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^{-1/2} (\mu_{1,n_1} - \mu_{2,n_2}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tan_\mu)$, *as* $\min(n_i) \to \infty$, *where* $\mu = \mu_1 = \mu_2$,
4. $t_{n_1,n_2} = (\mu_{1,n_1} - \mu_{2,n_2})^T (\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^{-1} (\mu_{1,n_1} - \mu_{2,n_2}) \xrightarrow{\mathcal{D}} \chi^2_m$, *as* $\min(n_i) \to \infty$.

Notice that the statistics t_{n_1,n_2} and their limit distribution do not depend on μ.

Remark 4 This time the $(E_k - \tan_{\mu_{i,n_i}})$-terms play a more essential role, namely in avoiding the otherwise cumbersome Moore inverse of $(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})$, since then S_1^2/n_1 and S_2^2/n_2 would be supported by different tangent spaces and their sum could have rank larger than m.

Proof of Theorem 4. Statement 1 follows from well-known theorems for convergence in probability (see, for example, Fuller (1976), Ch. 5.1). Statement 2 corresponds to Theorem 3, Statement 2. By straightforward calculations we obtain, with $c_{n_1,n_2} = \sqrt{\frac{1/n_1}{1/n_1 + 1/n_2}}$,

$$\begin{aligned} T_{n_1,n_2} &= (c^2_{n_1,n_2} S_1^2 + (1 - c^2_{n_1,n_2}) S_2^2)^{-1/2} \times \\ &\quad [c_{n_1,n_2} G^{-1}_{1,n_1} \tilde{V}^{1/2}_{1,n_1} \xi_{1,n_1} - (1 - c^2_{n_1,n_2})^{1/2} G^{-1}_{2,n_2} \tilde{V}^{1/2}_{2,n_2} \xi_{2,n_2}]. \end{aligned} \tag{4}$$

We will show that for any collection of numbers $0 \le c_{n_1,n_2} \le 1$

$$T_{n_1,n_2} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tan_\mu), \text{ as } \min(n_1, n_2) \to \infty. \tag{5}$$

Lemma 3 *Be given a directed set A and a collection of random variables $\{X_\alpha\}_{\alpha\in A}$ with values in Ω_0. Be given some compact set K, and a measurable function $g: K\times\Omega_0\to\Omega$ and a probability distribution Q on $(\Omega,\Im)$, $\Im$ is a σ-field in Ω, with the property that for any converging sequence in K, $c_m\to c$, as $m\to\infty$ and increasing sequence $\alpha_m\in A$ we have $g(c_m,X_{\alpha_m})\xrightarrow{\mathcal{D}}Q$. Then for any $\{c_\alpha\}_{\alpha\in A}$ in K, we have $g(c_\alpha,X_\alpha)\xrightarrow{\mathcal{D}}Q$.*

Proof. Suppose given c_α and some bounded continuous function f such that $E(f(g(c_\alpha,X_\alpha)))$ does not converge to $E_Q(f)$. Then we can certainly find an increasing subsequence α_m for which $E(f(g(c_{\alpha_m},X_{\alpha_m})))$ either converges to some other value than $E_Q(f)$ or diverges to ∞ or $-\infty$. But from the compactness of K we could find a converging subsequence of the c_{α_m}, and this would have exactly the same property. But this would contradict the hypothesis. □

Let us apply this Lemma in the proof of Theorem 4. Let A correspond to the set $\{1,2,3,...\}\times\{1,2,3,...\}$, $X_{n_1,n_2}=(S_1^2,S_2^2,G_{1,n_1},G_{2,n_2},\tilde V_{1,n_1},\tilde V_{2,n_2},\xi_{1,n_1},\xi_{2,n_2})$. Let $K=[0,1]$ and

$$g(c,X_{n_1,n_2})=(c^2S_1^2+(1-c^2)S_2^2)^{-1/2}[cG_{1,n_1}^{-1}\tilde V_{1,n_1}^{1/2}\xi_{1,n_1}-(1-c^2)^{1/2}G_{2,n_2}^{-1}\tilde V_{2,n_2}^{1/2}\xi_{2,n_2}]$$

Then

$$g(c_{n_1,n_2},X_{n_1,n_2})=T_{n_1,n_2}=(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2})^{-1/2}(\mu_{1,n_1}-\mu_{2,n_2}).$$

From well known theorems on convergence in distribution (see, for example, Fuller (1976), Theorems 5.2.5, 5.2.6) it follows that for c_{n_1,n_2} such that $c_{n_1,n_2}\to c$ for $\min(n_1,n_2)\to\infty$,

$$(c_{n_1,n_2},X_{n_1,n_2})\xrightarrow{\mathcal{D}}c\times\sigma_1^2\times\sigma_2^2\times G_1\times G_2\times\tilde V_1\times\tilde V_2\times\mathcal{N}(0,\tan_\mu)\times\mathcal{N}(0,\tan_\mu)$$

(the c_{n_1,n_2} being considered as deterministic random variables). Then, g being continuous along the support of the limit distribution and σ_1^2 and σ_2^2 positive definite, we obtain as a consequence of the Continuous Mapping Theorem (CMT) (see Billingsley (1968), Ch.1, Theorem 5.1),

$$g(c_{n_1,n_2},X_{n_1,n_2})\xrightarrow{\mathcal{D}}(c^2\sigma_1^2+(1-c^2)\sigma_2^2)^{-1/2}(cG_1^{-1}\tilde V_1^{1/2}Z_1-\sqrt{1-c^2}G_2^{-1}\tilde V_2^{1/2}Z_2),\qquad(6)$$

where Z_1 and Z_2 are independent, identically $\mathcal{N}(0,\tan_\mu)$ distributed.

Notice that $Z=(cG_1^{-1}\tilde V_1^{1/2}Z_1-\sqrt{1-c^2}G_2^{-1}\tilde V_2^{1/2}Z_2)$ is normally distributed with mean 0 and covariance matrix

$$EZZ^T=c^2G_1^{-1}\tilde V_1^{1/2}\tan_\mu\tilde V_1^{1/2}G_1^{-1}+(1-c^2)G_2^{-1}\tilde V_2^{1/2}\tan_\mu\tilde V_2^{1/2}G_2^{-1}$$
$$=c^2\sigma_1^2\tan_\mu+(1-c^2)\sigma_2^2\tan_\mu.$$

The last equality holds because $\tan_\mu$ commutes with the matrices $\tilde V_i$, G_i and $\sigma_i^2=G_i^{-1}\tilde V_iG_i^{-1}$. The limit statistic in (6) is also normally distributed with covariance matrix

$$E[(c^2\sigma_1^2+(1-c^2)\sigma_2^2)^{-1/2}ZZ^T(c^2\sigma_1^2+(1-c^2)\sigma_2^2)^{-1/2}]=$$
$$(c^2\sigma_1^2+(1-c^2)\sigma_2^2)^{-1/2}(c^2\sigma_1^2\tan_\mu+(1-c^2)\sigma_2^2\tan_\mu)(c^2\sigma_1^2+(1-c^2)\sigma_2^2)^{-1/2}=\tan_\mu\qquad(7)$$

Thus the underlying distribution Q in the Lemma corresponds to $\mathcal{N}(0, \tan_\mu)$. From the Lemma it follows that

$$T_{n_1,n_2} = g(c_{n_1,n_2}, X_{n_1,n_2}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tan_\mu),$$

for any c_{n_1,n_2}, as $\min(n_1, n_2) \to \infty$. Thus Statement 3 of the Theorem is proved. Statement 4 follows from Statement 3. $\square$

6 Experience with simulations

In this section we report on our experience with simulation of observations on Stiefel manifold $\mathbf{V}_{p,r}$, $r \le p$ (Section 4.1). We will use von Mises-Fisher distributed observations, given by its mean location $\mu \in V_{p,r}$ and a concentration parameter $\kappa > 0$, leading to densities $f_{\mu,\kappa}$ defined with respect to the uniform density on $V_{p,r}$, up to a proportionality factor C_κ, by

$$f_{\mu,\kappa}(X) = C_\kappa \exp(-\kappa \mathrm{Tr}\,(X-\mu)^T (X-\mu)).$$

We were interested in the performance of our χ^2-type procedures in the one-sample and two-sample cases. In spite of the seeming complexity in the formulation of the procedures everything can be implemented easily and performs well. We would like to indicate the three non-obvious steps one has to go through: simulation of uniform distribution on Stiefel manifold, simulation of von Mises-Fisher distribution on Stiefel manifold and programming Φ in (3) involving the inverse of the square root of positive definite matrices.

The method for generating uniform variates on Stiefel manifolds is to characterize it as the distribution which is invariant under the left action by $O(p)$. This property is shared by the multivariate standard normal distributions on $\mathrm{Mat}_{p,r}$ and preserved by the mapping Φ (cf. Chikuse (1990b), James (1954), Watson (1983b)).

The von Mises-Fisher distribution is simulated by the acceptance-rejection (AR) method (see Best and Fisher (1979), Johnson (1987), Jupp and Mardia (1989) and others). Let us notice that the efficient method given in Best and Fisher (1979) is not applicable in our case, mainly by the lack of symmetry for the von Mises-Fisher distributions. No attempt has been made on improving the AR-method in speed.

The programming can be done in many packages like MATLAB, where many operations needed are built in. We have tested for the χ^2_m distribution of the χ^2-type statistics of Theorem 3 (Statement 3) and Theorem 4 (Statement 4) in the H_0–situation for the special case considered in vectorcardiography, namely Stiefel manifold $\mathbf{V}_{3,2}$, where we restricted ourselves to the von Mises-Fisher distribution with concentration parameter $\kappa = 1$ for the first sample and $\kappa = 1.25$ for the second sample. We have observed that even for sample size $n = 25, m = 25$ asymptotic theory is perfectly adequate.

Table 1 summerizes the result of 1000 repetitions of the experiment of simulation of samples of observations on $V_{3,2}$ of size $n = 25, 50, 100$. One can see that the KS statistics don't exceed the critical 95% level equal to 1.36 if $n > 50$ for the one sample case and if $n = m \ge 25$ for the two sample case. The phenomenon of better convergence to the limit distribution in the two sample case should not be surprising, and may be explained by the double information that is provided. Figures 1, 2 and 3 are complementing the KS statistic,

Table 1. The value of sample mean and variance
of χ^2 - type statistics from Theorem 3(3)
and Theorem 4(4) and Kolmogorov - Smirnov statistic
for 1000 repetitions of the experiment

One - sample test				
n	25	50	100	χ_3^2
mean	3.5515	3.3447	3.0964	3
variance	9.4978	8.9439	6.5671	6
KS	2.552	1.3640	0.7914	random
Two - sample test				
n, m	25	50	100	χ_3^2
mean	3.0962	3.1369	3.0996	3
variance	7.3365	6.9178	6.3423	6
KS	0.6845	0.7511	0.8003	random

by giving the graph of $n^{1/2}(\hat{F}(x) - F(x))$ against $F(x)$, where $\hat{F}$ is the empirical distribution function of our testing statistic and F is the distribution function of χ_3^2. The outer dashed lines are at the levels ± 1.36.

All simulations and testing procedures were done using MATLAB programming language, version 4.2c.1, running on SUN - SS10 sparcstation computers (MathWorks (1995)). In order to enable duplication of the obtained random numbers and graphs, the MATLAB code will follow. First there is a common prologue.

```
global p r ;p=3;r=2; % whatever you may like, but r<=p.
m=r*p-r*(r+1)/2;     % dimension of Stiefel manifold Vpr
global mu; kappa=1; mu=eye(p,r); % parms von Mises-Fisher distribution

thesamples=[];
ss=1000                 % number of replications of experiment
samplesize=[25 50 100] % size of experiments
```

This is the part pertinent to the one sample case.

```
%One-sample tests
randn('seed',232080324);rand ('seed',931316785)
for n=samplesize,

smpl=zeros(1,ss);
for q=(1:ss),

z1=0; z2=0; for i=1:n, x=rvmfst(kappa); z1=z1+x; z2=z2+x(:)*x(:)'; end
Xbarn=z1/n; Vn=(z2-n*Xbarn(:)*Xbarn(:)')/n;
mun=phi(Xbarn); % empirical mean location
d=mun-mu;
```

Figure 1: Plot of $n^{1/2}$-deviations of empirical and theoretical distribution functions for the χ^2-type statistic. One sample test, $n = 50$.

Figure 2: Plot of $n^{1/2}$-deviations of empirical and theoretical distribution functions for the χ^2-type statistic. One sample test, $n = 100$.

Figure 3: Plot of $n^{1/2}$-deviations of empirical and theoretical distribution functions for the χ^2-type statistic. Two sample test, $n = 50, m = 50$.

```
Mn=d-weing(tang(d,mun),Xbarn-mun,mun);

T=tangm(mun); %produce matrix corresponding to tang( ,mun)

Vntwiddle=T*Vn*T+eye(p*r)-T;
Vntwiddleminus1=Vntwiddle∧(-1);
smpl(q)=n*Mn(:)'*Vntwiddleminus1*Mn(:);

end
thesamples=[thesamples;smpl];

end % end of one-sample tests
```

Now each of the three samples can be studied as follows:

```
for i=1:3
disp(['n=' int2str(samplesize(i)) '; replications ' int2str(ss)])
smpl=thesamples(i,:)
z=mean(smpl)                       % expected m
var=(smpl-z)*(smpl-z)'/(ss-1)      % expected 2*m
datasort=sort(smpl);
pchisq=gammainc(datasort/2,m/2);Y=(0:(ss-1))/ss;
KS=sqrt(ss)*max(abs(pchisq-Y))     % Kolmogorov-Smirnov statistic
```

```
% deviation-plot
crit=1.36;
plot(pchisq,sqrt(ss)*(pchisq-Y));hold on;axis([0 1 -1.5 1.5]);
plot([0 1],[crit crit],'--');plot([0 1],[-crit -crit],'--');
plot([0 1],[0 0],'--');hold off

end % end of one-sample tests
```

The two-sample code is as follows:

```
%Two-sample tests
randn('seed',322080324);rand('seed',391316785);
kappa1=kappa,kappa2=kappa*10/8
for n=samplesize,n1=n;n2=n;

smpl=zeros(1,ss);
for q=1:ss,
z1=0; z2=0; for i=1:n1, x=rvmfst(kappa1); z1=z1+x; z2=z2+x(:)*x(:)'; end
Xbarn1=z1/n1; Vn1=(z2-n1*Xbarn1(:)*Xbarn1(:)')/n1;
mun1=phi(Xbarn1); % empirical mean location

A1=weingm(Xbarn1-mun1,mun1);
B1=(eye(p*r)-A1)∧(-1);
T1=tangm(mun1);
%produce matrix S∧2_1
S1sq=B1*T1*Vn1*T1*B1+eye(p*r)-T1;

z1=0; z2=0; for i=1:n2, x=rvmfst(kappa2); z1=z1+x; z2=z2+x(:)*x(:)'; end
Xbarn2=z1/n2; Vn2=(z2-n2*Xbarn2(:)*Xbarn2(:)')/n2;
mun2=phi(Xbarn2); % empirical mean location

A2=weingm(Xbarn2-mun2,mun2);
B2=(eye(p*r)-A2)∧(-1);
T2=tangm(mun2);
%produce matrix S∧2_2
S2sq=B2*T2*Vn2*T2*B2+eye(p*r)-T2;

%The test statistic
Ssq=(S1sq/n1+S2sq/n2);
t=(sqrtm(Ssq))∧(-1)*(mun1(:)-mun2(:));
T=t'*t;

smpl(q)=T;
end
```

```
thesamples=[thesamples;smpl];

end % end of two-sample tests
```

The three samples can be studied exactly as in the case of the one sample simulation. And then there are the MATLAB functions for these programs:

```
% put next lines in file phi.m
function y=phi(x)
%phi            orthogonal projection on Stiefel manifold
y=x/sqrtm(x'*x);

% put next lines in file runifst.m
function y=runifst
%runifst   random uniformly distributed Stiefel matrix.
global p r
y=phi(randn(p,r));

% put next lines in file rvmfst.m
function y=rvmfst(kappa)
%rvmfst   random von Mises-Fisher distributed Stiefel matrix.
global mu % and p r via runifst
z=0;t=1;
while z<t,
rv=runifst; z=exp(-kappa*(rv(:)-mu(:))'*(rv(:)-mu(:)));
t=rand;
end;
y=rv;

% put next lines in file tang.m
function y=tang(U,mu)
%tang   projection at tangent space at mu, only if f(mu)=eye(size(mu,2))
%       (defining equation for Stiefel manifold) where f(mu)=mu'*mu.
%       orthogonal projection:  y*y=y; y'=y;
%       onto tangent space:  df(mu)*y=0;
%               rank(y)=size(mu,1)*size(mu,2)-size(mu,2)*(size(mu,2)+1)/2
y=U-(1/2)*mu*(mu'*U+U'*mu);

% put next lines in file tangm.m
function y=tangm(mu)
% tangm   Produces matrix of the linear mapping tang(.,mu)
[p r]=size(mu);
y=eye(p*r);
for i=(1:p*r), u=tang(reshape(y(:,i),p,r),mu);y(:,i)=u(:);end

% put next lines in file weing.m
```

```
function y=weing(W,nmu,mu)
%weing       Weingarten map at M wrt normal direction N
y=-W*mu'*nmu+(1/2)*mu*(mu'*W*mu'*nmu+nmu'*mu*W'*mu);

% put next lines in file weingm.m
function y=weingm(nmu,mu)
% weingm   Produces matrix of the linear mapping weingm(.,nmu,mu)
[p r]=size(mu);
y=eye(p*r);
for i=(1:p*r), u=tang(reshape(y(:,i),p,r),mun);
u=weing(reshape(y(:,i),p,r),mu);
y(:,i)=u(:);end
```

References

[1] Best, D.J. and Fisher, N.I. (1979). Efficient simulation of the von Mises distribution. *Applied Statistics*, **24**, 152-157.

[2] Bickel, P.J. and Doksum, K.A. (1977). *Mathematical Statistics*, Holden-Day, San Fransisco.

[3] Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[4] Chikuse, Y. (1990a). Distributions of orientations on Stiefel manifolds. *J. Multivariate Analysis*, **33**, 247 - 264.

[5] Chikuse, Y. (1990b). The matrix angular central Gaussian distribution. *J. Multivariate Analysis*, **33**, 265 - 274.

[6] Downs, T. D. (1972). Orientation statistics. *Biometrika*, **59**, 665 - 676.

[7] Fisher, N.I., Lewis, T., Embleton, B.J.J. (1987). *Statistical Analysis of Spherical Data.* Cambridge University Press, London.

[8] Fuller, W.A. (1976). *Introduction to Statistical Time Series.* Wiley, New York.

[9] Hendriks, H. (1990), Sur le cut-locus d'une sous-variété de l'espace euclidien, *C.R. Acad. Sci. Paris. Sér. I*, **311**, 637-639.

[10] Hendriks, H. (1991), A Cramér-Rao type lower bound for estimators with values in a manifold, *J. Multivariate Analysis*, **38**, 245-261.

[11] Hendriks, H., Janssen, J. & Ruymgaart, F. (1992), A Cramér-Rao type inequality for random variables in Euclidean manifolds. *Sankhyā–Ser.* A **54**, 387-401.

[12] Hendriks, H. (1992), Sur le cut-locus d'une sous-variété de l'espace euclidien. Négligeabilité, *C.R. Acad. Sci. Paris, Sér I*, **315**, 1275-1277.

[13] Hendriks, H., Landsman, Z. & Ruymgaart, F. (1996), Asymptotic behavior of sample mean direction for spheres, *J. Multivariate Analysis*, **59**, 141-152.

[14] Hendriks, H., Landsman, Z (1996a), Asymptotic behavior of sample mean location for manifolds. *Stat. Probab. Letters*, **26**, 169-178

[15] Hendriks, H., Landsman, Z. (1996b), Asymptotic tests for mean location on manifolds. *C.R.Acad Sci.Paris*, Série A, **322**, 773-778.

[16] James, A.T. (1954), Normal multivariate analysis and the orthogonal group. *Annals of Mathematical Statistics*, **25**, 40-75.

[17] Johnson, M.E. (1987). *Multivariate Statistical Simulation*, Wiley, New York.

[18] Jupp, P.E. and Mardia, K.V. (1979), Maximum likelihood estimation for the matrix von Mises - Fisher and Bingham distributions. *Ann. Statist.* **7**, 599 - 606.

[19] Jupp, P.E. and Mardia, K.V. (1989), A unified view of the theory of directional statistics, 1975 - 1988. *International Statistical Review*, **57**, 3, 261 - 294.

[20] Khatri, C.G. and Mardia, K. V. (1977), The von Mises - Fisher matrix distribution in orientation statistics, J.*R. Statist. Soc.* B, **39**, 1, 95 - 106.

[21] Kobayashi S. & Nomizu, K. (1969). *Foundations of Differential Geometry*, Vol.2. Interscience, New York.

[22] Lancaster, P., (1969). *Theory of Matrices*. Academic Press, New York - London.

[23] MathWorks (1995) *Matlab User's Guide For Unix Workstations*, MathWorks, IncSouth Natick, MA.

[24] Mardia, K. V. (1975a). *Statistics of Directional Data*. Academic Press, London.

[25] Mardia, K. V. (1975b). Statistics of directional data, *J.R. Statist. Soc.* B, **37**, 3, 349 - 371.

[26] Mardia, K. V. and Khatri, C.G. (1977), Uniform distribution on a Stiefel manifold *J. Multivariate Analysis*, **7**, 468 - 473.

[27] Mardia, K. V. (1988), Directional data analysis: an overview, *Journal of Applied Statistics*, **15**, 115-122.

[28] Prentice, M.J. (1986), Orientation statistics without parametric assumptions. *J.R. Statist. Soc.* B, **48**, 214 - 222 .

[29] Prentice, M. J. (1989), Spherical regression on matched pairs of orientation statistics. *J. Roy. Statist. Soc.* B, **51**, 2, 241–248.

[30] Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley, New York.

[31] Thom, R. (1972), Sur le cut-locus d'une variété plongée, *J. Diff. Geom.*, **6**, 577-586.

[32] Watson, G. S. (1983a). *Statistics on Spheres*. Wiley, New York.

[33] Watson, G. S. (1983b). Limit theorems in high dimensional spheres and Stiefel manifolds. In *Studies in Econometrics, Time Series, and Multivariate Statistics*, Ed. S. Karlin, T. Amemiya, L. Goodman, Academic Press, New York.

Department of Mathematics, University of Nijmegen,
Toernooiveld, 6525 ED Nijmegen, Pays Bas.
e-mail address: hhendr@sci.kun.nl
Department of Statistics, University of Haifa,
Mount Carmel, Haifa 31905, Israel.
e-mail address: landsman@stat.haifa.ac.il