Comparing Random Variables on Manifolds

Nikolay H. Balov *

Abstract

We are interested in comparing probability distributions defined on Riemannian manifolds. The traditional approach to study a distribution relies on locating its mean point and finding the dispersion about that point. On a general manifold however, even if two distributions are sufficiently concentrated and have unique means, a comparison of their covariances is not possible due to the difference in local parametrizations. To circumvent the problem we associate a covariance field with each distribution and compare them at common points by applying a similarity invariant function on their representing matrices. In this way we are able to define distances between distributions. We also propose new approach for interpolating discrete distributions and derive some criteria that assure consistent results. Finally, we illustrate with some experimental results on the unit 2-sphere.

Key Words: non-parametric, manifolds, covariances, interpolation

1. Introduction

The problem of comparing distributions defined on non-Euclidean spaces or to be more specific, on Riemannian manifolds, becomes increasingly important. A typical example of non-trivial manifold is the unit 2-sphere $\mathbb{S}^2$, which is the domain of our experiments in this work. Consequently, our study has as main application, but not limited to, problems from directional statistics, a branch of statistics dealing with directions and rotations in $\mathbb{R}^3$.

Pioneers in the field are Fisher, R.A.(1953) and von Mises. In recent years directional statistics proved to be useful in variety of disciplines like shape analysis [10], geology and crystallography [9]. Most of the practitioners in these fields use parametric distributions to model directional data, like von Mises-Fisher distribution on the circle and Fisher-Bingham-Kent(FBK) distributions on the 2-sphere. For more details see Kent, J. (1982).

There are application areas however, where parametric models are insufficient. A recent example is provided by medical imaging community. In a new technique based on MRI and called High Angular Resolution Diffusion Imaging (HARDI), the data is represented by Orientation Distribution Functions (ODFs) which are nothing but discrete distributions on the unit 2-sphere. These distributions by their nature are multi-modal - not concentrated about a particular direction. They do not follow a parametric model and even if they do the eventual model would be too complicated to be efficient. Consequently, a non-parametric approach is more natural in processing ODFs.

In analysis of HARDI data researchers first have to solve the problem of registration between different volumes of ODFs, corresponding to the images of different subjects. For this purpose they need models and algorithms for interpolation between ODFs. There are no many choices for interpolation procedure beyond the simplest linear one. A recent alternative, using the square root representation of probability mass functions, was proposed by Srivastava(2007) and implemented in [5]. No existing solution though respects the geometry of the underlying domain.

It is the main subject of this paper to draw some new directions for searching of possible solutions - approaches that address the non-Euclidean nature of the random variables and provide adequate solutions.

What we propose basically is a generalization of the classical concept of covariance of distribution. We allow covariance to be defined with respect to any point of distribution domain and by doing so we try to workaround the problem of finding the mean point, which might not exist or be ambiguous. Also, since compact manifolds like $\mathbb{S}^2$ do not admit global parametrizations, we pay special attention to use the correct mathematical tool for describing the covariance. We not only point out to the well known fact that covariance can be viewed as a bi-linear operator and thus defined as a tensor, but specify the exact variety of this tensor. It is important to make a distinction between covariance tensor and metric tensor on manifold. A central observation in our approach is that at any point of the domain, the product of the metric and covariance tensors is a linear operator on the respected tangent space. We call it *covariance operator*. Collectively they form a field of operators. By applying appropriate functions, so called similarity invariants, on a covariance operator field we are able to define a representation of the underlying distribution. These distribution representations can be used for the purpose of comparison and interpolation.

Although in all our experiments we stay on the unit sphere , the theoretical framework still holds on a general Riemannian manifold and this is one of its main advantages.

*Florida State University, Department of Statistics, balov@stat.fsu.edu

2. Covariance fields

2.1 Random variables on manifold

Let M be a Riemannian n-manifold, $q \in M$ and let Exp_q be the *exponential map* at q, $Exp_q : M_q \to M$. It sends a tangent vector at q to a point on the manifold along a geodesic through q. There is a maximal open set $U(q)$ in M_p containing the origin, where Exp_q is a diffeomorphism. Then the set $\mathcal{U}(q) = Exp_q(U(q))$ is called maximal normal neighborhood of q. On this normal neighborhood the exponential map is invertible and let

$$Log_q = Exp_q^{-1} : \mathcal{U}(q) \to M_p$$

be its inverse, the so called *log-map*. Log_q is diffeomorphism on $\mathcal{U}(q)$. We adopt the notation $\overrightarrow{qp} = Log_q p$ in analogy to the Euclidean case, $M = \mathbb{R}^n$, where $Log_q p = p - q = \overrightarrow{qp}$.

In particular, on $M = \mathbb{S}^n$ the log-map has a closed-form expression

$$\overrightarrow{qp} = \frac{cos^{-1} < p, q >}{(1 - < p, q >^2)^{1/2}}(p - < p, q > q), \tag{1}$$

which greatly simplifies all metric related operations on the unit sphere.

The Borel sets on M generated by the open sets on M form a σ-algebra $\mathcal{A}(M)$ on M. Any Riemannian manifold has a natural measure $\mathcal{V}$ on $\mathcal{A}(M)$, called *volume measure*. In local coordinates x it is given by $dV(x) = \sqrt{|G_x|}dx$, where G_x is the matrix representation of the metric tensor, $|G_x|$ is its determinant and dx is the Lebesgue measure in $\mathbb{R}^n$. For example, $\mathbb{S}^2$ parametrized with geographical coordinates (θ, ϕ) has a volume form $V(\theta, \phi) = \cos(\theta)d\theta d\phi$.

A random variable X on M is any measurable function from a probability space $(\Omega, \mathcal{B}, \mathcal{P})$ to $(M, \mathcal{A}(M), \mathcal{V})$. The distribution function F of X is defined as $F(A) = \mathcal{P}(X^{-1}(A)), A \in \mathcal{A}(M)$. If F can be represented by

$$F(A) = \int_A f(p)dV(p), \forall A \in \mathcal{A}(M),$$

for almost everywhere continuous (w.r.t. $\mathcal{V}$) function f, then F is said to be absolute continuous (w.r.t. $\mathcal{V}$) and f is its density (*pdf*).

2.2 Intrinsic and Extrinsic mean and covariance

Let (M, ρ) be a metric space. The *Fréchet mean* set of a distribution F is the set of minimizer's of $Q(q) = \int \rho^2(q, p)dF(p)$. It was introduced by Frechet (1948). If M is a Riemannian manifold M with metric structure g, then the intrinsic mean of F, is the Frechet mean of (M, d_g), where d_g is the geodesic distance. Karcher(1977) considered the intrinsic mean on M and gave conditions for its existence and uniqueness. An alternative to intrinsic mean is the extrinsic one, which is obtained by embedding M into a higher dimensional Euclidean space. We point to the influential paper of Bhattacharya R. and Patrangenaru, V. (2003) where the properties of extrinsic and intrinsic means and their relation and asymptotic properties are considered in details.

Once a mean point (intrinsic or extrinsic) is specified, the covariance can be defined as usual after fixing a coordinate system about that point.

To compare two distributions one may first look at their intrinsic means. If they differ, the distributions differ, otherwise one may further compare their covariances at the common mean point. This approach however suffers from at least two drawbacks. First, if the population mean set is large, then the finite sample intrinsic mean will have substantial variance. That will diminish the power of any test for equality of means and more importantly, will inevitably require comparing covariances at different points. Second, the intrinsic mean, provided it exists and it is unique, and the covariance alone do not specify completely the distribution.

Thus for solving the problem of comparing distributions, we need a more informative structure that completely represents distributions and that is defined in coordinates free manner for seamless manipulation and analysis.

2.3 Covariance operators

Many parametric families of distributions can be defined as functions on linear operators. Consider for example the standard normal distribution in $\mathbb{R}^n$ with density $f(x) \propto \exp(-\frac{1}{2}||x - \mu||^2)$, where $\mu \in \mathbb{R}^n$ is its mean. Since $||x - \mu||^2 = tr((x - \mu)(x - \mu)')$ and the matrix $L(x) = (x - \mu)(x - \mu)'$ defines a linear operator

$$L(x)(u, v) = u'L(x)v = [u'(x - \mu)][(x - \mu))'v], u, v \in \mathbb{R}^n,$$

we can express the density by $f(x) \propto h(L(x))$, $h(T) := \exp(-\frac{1}{2}tr(T))$.

Let M be a Riemannain manifold with metric G. If $q \in M$, then $G(q)$ is a co-variant 2-tensor at M_q, while the quantity $(\overrightarrow{qp})(\overrightarrow{qp})'$ is a contra-variant 2-tensor at M_q. The contraction of their tensor product, $G(q)(\overrightarrow{qp})(\overrightarrow{qp})'$, is a

(1,1)-tensor, or equivalently, a linear operator at M_q. For a distribution F on M, we define a linear operator at M_q by taking the expectation of $G(q)(\overrightarrow{qp})(\overrightarrow{qp})'$ with respect to F.

From now on we will use the standard notation $T^2(M_q)$ for co-variant 2-tensors on M_q, $T_2(M_q)$ for contra-variant 2-tensors on M_q and $T_1^1(M_q)$ for bi-linear operators on M_q.

Definition 1 *Let $r : \mathbb{R}^+ \to \mathbb{R}^+$ be a continuous function. Covariance of distribution F on M at point $q \in M$ is defined by*

$$\Sigma(q) = \int_{\mathcal{U}(q)} (\overrightarrow{qp})(\overrightarrow{qp})' r(||\overrightarrow{qp}||) dF(p) \tag{2}$$

and $\Sigma : q \mapsto \Sigma(q) \in T_2(M_q)$ is called covariance field of F.

With $r = 1$ we obtain the generic covariance field associated with F and this is the default choice.

As noted above, $G(q)\Sigma(q)$ is a linear operator on M_q, which we call *covariance operator*. Hence, $G\Sigma$ is a field of linear operators on M. With respect to a coordinate system x at q, $G(q)\Sigma(q)$ is represented by a symmetric and positive definite matrix $G_x\Sigma_x$, where G_x and Σ_x are the representations of $G(q)$ and $\Sigma(q)$ respectively. In other words, $G\Sigma$ is a field of symmetric and positive definite operators on M.

If $v \in M_q$ has components v_x with respect to x, we define $(G(q)\Sigma(q))v := \Sigma_x G_x v_x$ and $< v, (G(q)\Sigma(q))v >:= v_x' G_x \Sigma_x G_x v_x$. One can check that indeed the last quantity is invariant to coordinate change at q.

It is worth to mention that for a covariance field Σ on M, Σ^{-1} is also symmetric and positive definite and when it is differentiable, Σ^{-1} introduce a new Riemannian metric on M. Moreover, if Σ_1 and Σ_2 are two covariance fields on M, then $\Sigma_1\Sigma_2^{-1}$ is a field of linear operators, i.e. for any $q \in M$, $\Sigma_1(q)\Sigma_2^{-1}(q) \in T_1^1(M_q)$.

On a complete Riemannian manifold, the problem of minimizing the trace of the default covariance field is equivalent to the problem of finding the intrinsic mean μ of F, i.e.

$$\mu = argmin_{q \in M} \{ \int_{\mathcal{U}(q)} tr(G(q)(\overrightarrow{qp})(\overrightarrow{qp})') dF(p) = \int_M d_g^2(q,p) dF(p) \}.$$

See [2] for more comprehensive introduction to covariance fields.

2.4 Similarity invariants

Let Sym_n^+ denote the space of symmetric and positive definite matrices. Since this is the representation domain of covariance operators it is of obvious importance for us. Sym_n^+ attracted the attention of many researchers in the recent years due to its manifold nature that accepts convenient metric structures and consequently, the variety of application opportunities it provides. For the purposes of Diffusion Tensor Imaging, Fletcher, P. T., Joshi, S., (2007) and Pennec. X., Fillard, P., Ayache, N (2006) proposed the use of *affine invariant* distance, while Arsigny,V., Fillard, P., Pennec X., and Ayache, N. (2007) proposed the so called log-Euclidean distance. A good survey of the available distances and estimators in Sym_n^+ along with new ones is provided by Dryden, I., etc. (2008). We aim a more general treatment of Sym_n^+ and instead of dealing with specific matrix functions we define a whole class of invariants. What particular member of this class should be used is an application specific choice.

Two matrices $A, B \in Sym_n^+$ are said to be similar if

$$A = X^{-1}BX, \text{ for } X \in GL_n.$$

Matrix representations of linear operators are similar and thus, this fact holds for the representations of $G\Sigma$ and $\Sigma_1\Sigma_2^{-1}$. Next we define an important class of functions that respect similarity.

Definition 2 *A similarity invariant function on Sym_n^+ is any continuous bi-variate h that satisfies*

(i) $h(AXA', AYA') = h(X,Y)$, $\forall X, Y \in Sym_n^+$ *and* $A \in GL_n$.

It is a non-negative with a unique root if

(ii) $h(X,Y) \geq 0$, $\forall X, Y \in Sym_n^+$ *and* $h(X,Y) = 0 \iff X = Y$.

Moreover, h is called similarity invariant distance, if in addition to (i) and (ii) also satisfies

(iii) $h(X,Y) + h(Y,Z) \geq h(X,Z)$, $\forall X, Y, Z \in Sym_n^+$.

Below we list several examples of similarity invariant function we use in our experiments.

1. For a fixed $Z \in Sym_n^+$, the similarity invariant

$$h_{trdif}(X, Y; Z) = |(tr(Z^{-1}X - Z^{-1}Y)|,$$

satisfies (iii) but not (ii). Default choice will be $Z = G^{-1}$, the inverse of the metric tensor representation.

2. The second one is sometimes referred as *affine-invariant distance* in Sym_2^+, see for example [11], [4], [7] and [12], and it is defined by
$$h_{trln2}(X,Y) = \{tr(ln^2(XY^{-1}))\}^{1/2}, X, Y \in Sym_2^+.$$
Actually, h_{trln2} is not a unique choice for a distance in Sym_2^+.

3. Log-likelihood function gives us another choice for h,
$$h_{lik}(X,Y) = tr(XY^{-1}) - ln|XY^{-1}| - n.$$
It satisfies (i) and (ii) but it fails to satisfy the triangular inequality.

The concept of covariance fields can be used for measuring the difference between distributions on M. Let f and g be two densities on M and $\Sigma[f]$ and $\Sigma[g]$ be their respected covariance fields.

For a non-negative $h \in \mathcal{SIM}(n)$ we define
$$d_h(f,g) := \int_M h(\Sigma[f](p), \Sigma[g](p))dV(p). \tag{3}$$
When M is a compact, the above integral is well defined and finite. Moreover, if $h(X,Y)$ is a distance function on Sym_n^+, then d_h will be a distance in the space of densities on M.

Equation (3) gives a very general but impractical way to compare distributions due to the fact that the integration domain is the whole manifold. For application purposes however, one may restrict to a smaller domain or perform the comparison on discrete set of points which are of particular interest.

3. Interpolation of discrete distributions on $\mathbb{S}^2$

Here we assume that distributions are defined on a common domain - a fixed set of points on the sphere. The approach we propose is first, to generate an interpolated field based on the covariance fields of the initial distributions and second, to find a probability mass function which covariance field is close to the interpolated one. Closeness is measured using a suitable similarity invariant function. Covariance fields are also considered discrete ones - they are defined on a finite set of observation points. With a fixed coordinate system at each observation point, not necessarily a global one, the covariance field is represented by a set of matrices. As always, we are going to use the tensor notation to guarantee a coordinate free approach.

Let $\{p_i\}_{i=1}^k$ and $\{q_i\}_{i=1}^k$ be two sets of k points on $\mathbb{S}^2$. The first set is the distribution domain. The second one is the observation set. Hereafter, a discrete mass function (pmf) is any k-vector f, such that $f = \{f_i = f(p_i) \geq 0\}_{i=1}^k$ and $\sum_{i=1}^k f_i = 1$. We write $f \in P_k^+$, where P_k^+ denotes the compact k-simplex.

The number of observation points may be in fact less than k, the size of the *pmfs*. However, with a smaller observation set one may lose the uniqueness and the continuity of an estimation. Particular geometric configurations also lead to the same result and one has to check carefully the consistency conditions corresponding to the problem.

The covariance field of $f \in P_k^+$ at q_j is defined as
$$\Sigma[f]_j := \Sigma[f](q_j) = \sum_{i=1}^k (\overrightarrow{q_j p_i})(\overrightarrow{q_j p_i})' r(||\overrightarrow{q_j p_i}||) f(p_i),$$
where $\overrightarrow{q_j p_i}$ is given by (1). We use either $r = 1$ or $r(t) = (1 - \frac{\pi}{2t})^2$. The second choice is known to be optimal on $\mathbb{S}^2$ in the class of functions $r_a(t) = (1 - \frac{a}{t})^2$ because it minimizes the maximum of $tr(G\Sigma(q))$ (see Lemma 3 in [2]).

Let f^s,s=1,...,m, be a collection of *pmfs* and $\{C_j^s = \Sigma[f^s]_j\}_{j=1}^k, s = 1, ..., m$, be their covariances. With a non-negative similarity invariant function h, we measure the difference between f and f^s
$$d_h(f, f^s) := \sum_{j=1}^k h(\Sigma[f]_j, C_j^s), s = 1, ..., m. \tag{4}$$

Let $\alpha \in P_m^+$, i.e. $\alpha = \{\alpha_s\}_{s=1}^m$, such that $\alpha_s \geq 0$ and $\sum_s \alpha_s = 1$. Then we define the functional
$$H(f; \alpha) := \sum_{s=1}^m \alpha_s d_h(f, f^s). \tag{5}$$

Finally, we formulate the following optimization problem: find a probability mass function $\hat{f}$ such that
$$\hat{f}(\alpha) = argmin_f H(f; \alpha). \tag{6}$$

Below we show some results regarding the consistency of the estimators (6).

Lemma 1 *Let $h \in \mathcal{SIM}(n)$, $\alpha^l \in P_M^+$ and $f^l \in P_k^+$. If $\alpha^l \to \alpha^0$ and $f^l \to f^0$ (in L_2 norm), then*

$$H(f^l, \alpha^l) \to H(f^0, \alpha^0).$$

Proof: We have that $\{\Sigma[f]_j\}_{j=1}^k$ are continuous in f (see Prop. 2 in [2]). Now observe that

$$||H(f^l;\alpha^l) - H(f^0;\alpha^0)|| \leq ||H(f^l;\alpha^l) - H(f^l;\alpha^0)|| + ||H(f^l;\alpha^0) - H(f^0;\alpha^0)||.$$

Since $H(f;\alpha^0)$ is continuous in f, the second term above goes to zero. The first term is bounded by

$$||H(f^l;\alpha^l) - H(f^l;\alpha^0)|| \leq ||\alpha^l - \alpha^0|| \max_{s,j,l} h(\Sigma[f^s]_j, C_j^l).$$

The sets $\{\Sigma[f]_j | f \in P_k^+\}$ are compact in Sym_n^+ and h is continuous, therefore $\max_{s,j,l} h(\Sigma[f^s]_j, C_j^l) = C < \infty$ and $H(f^l;\alpha^l) \to H(f^l;\alpha^0)$. $\square$

For a sequence α^l, define $\hat{f}^l = argmin_f H(f;\alpha^l)$. We have the following

Lemma 2 *If $h \in \mathcal{SIM}(n)$ and $\alpha^l \to \alpha^0$, then $H(\hat{f}^l, \alpha^l) \to H(\hat{f}^0, \alpha^0)$.*

Proof: Since P_k^+ is a compact, any sub-sequence of $\hat{f}^l$ has a point of convergence in P_k^+. Without loss of generality we may assume that $\hat{f}^l \to g \in P_k^+$. Accounting for the minimizing properties of $\hat{f}$ and applying lemma 1 we can write

$$H(\hat{f}^0, \alpha^l) \geq H(\hat{f}^l, \alpha^l) \to H(g, \alpha^0) \geq H(\hat{f}^0, \alpha^0).$$

The claim follows from the fact that $H(\hat{f}^0, \alpha^l) \to H(\hat{f}^0, \alpha^0)$. $\square$

Note that $H(\hat{f}^l, \alpha^l) \to H(\hat{f}^0, \alpha^0)$ is not enough to claim that $\hat{f}^l \to \hat{f}^0$. However, if $H(f;\alpha^0)$ has a well separated minimum at $\hat{f}^0$, then indeed $\hat{f}$ is continuous at α^0.

Another problem is how to find the global minimum $\hat{f}$ of $H(f;\alpha)$, provided it is unique. We know that the minimum is easily found in case of convex function H, by gradient descent algorithm for example. Moreover, the convexity of $H(f;\alpha^0)$ in P_k^+ guarantees the well separability of its minimum and that gives us the desired consistency.

Proposition 1 *If $\alpha^l \to \alpha^0$ and $h \in \mathcal{SIM}(n)$ is such that $H(f;\alpha^0)$ is convex in P_k^+, then $\hat{f}^l \to \hat{f}^0$.*

Proof: Suppose the contrary, that there exists $g \in P_k^+$, and sub sequence $\hat{f}^l \to g$, such that $||\hat{f}^0 - g|| > 0$. Then $H(g;\alpha^0) > H(\hat{f}^0;\alpha^0)$ by the separability of the minimum. But $H(\hat{f}^l;\alpha^l) \to H(g;\alpha^0)$ by lemma 1 and $H(\hat{f}^l;\alpha^l) \to H(\hat{f}^0;\alpha^0)$ by lemma 2, which imply $H(g;\alpha^0) = H(\hat{f}^0;\alpha^0)$. The contradiction shows that the assumption for g is false, which proves the claim. $\square$

3.1 Linear Interpolation

Here we consider the similarity invariant function $h^2_{trdif}(.,.;G^{-1})$. The corresponding optimization functional is

$$H_{trdif}(f,\alpha) = \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} tr^2(G(q_j)\Sigma[f]_j - G(q_j)C_j^s)$$

We use the default covariance, $r = 1$. Denote $a_{ij} = tr(G(q_j)(\overrightarrow{q_j p_i})(\overrightarrow{q_j p_i})') = d^2(q_j, p_i)$ and $c_j^s = tr(G(q_j)C_j^s)$, then

$$H_{trdif}(f,\alpha) = \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} (\sum_i a_{ij} f_i - c_j^s)^2.$$

We have

$$\frac{\partial H_{trdif}}{\partial f_i} = 2 \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} a_{ij} (\sum_l a_{lj} f_l - c_j^s).$$

The second partial derivatives are

$$\frac{\partial^2 H_{trdif}}{\partial f_i \partial f_l} = 2 \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} a_{ij} a_{lj}$$

Let $w = \{w_i\} \in \mathbb{R}^k$, then

$$\sum_{i,l} w_i w_l \frac{\partial^2 H_{trdif}}{\partial f_i \partial f_l} = 2 \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} (\sum_{i=1}^{k} w_i a_{ij})^2 \geq 0.$$

Therefore, if the matrix $A = \{a_{ij}\}_{i=1,j=1}^{k,k}$ is of full rank k, then H_{trdif} is convex in P_k^+. Moreover, the optimal solution of (6) satisfies $\sum_i a_{ij} f_i = \sum_{s=1}^{m} \alpha_s c_j^s$,
$j = 1, ..., k$, with a unique solution $\hat{f} = \sum_{s=1}^{m} \alpha_s f^s$, since for every s and j, $\sum_i a_{ij} f_i^s = c_j^s$.

Thus, we showed the following

Proposition 2 *If the matrix A has full rank, rank(A) = k, then the linear interpolation is the unique solution of the optimization problem (6) for* H_{trdif}.

3.2 Non-Linear Interpolations

Consider similarity invariant function h_{trln2} and corresponding optimization functional H_{trln2}

$$H_{trln2}(f; \alpha) = \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} tr(ln^2(\Sigma[f]_j (C_j^s)^{-1})).$$

The value of $H_{trln2}(f)$ is small when $G(q_j)\Sigma[f]_j$ is close to covariance operators $G(q_j)C_j^s$ for all j and s. This is a much stronger condition than the requirement for their traces to be close as in the problem of minimizing H_{trdif}. Consequently the minimum of $H_{trln2}(f)$, in general, will be strictly positive and the optimal pmf will be different from the linear interpolation.

Define the operators

$$Z_{ij}^s = (\overrightarrow{q_j p_i})(\overrightarrow{q_j p_i})'(1 - \frac{\pi}{2||\overrightarrow{q_j p_i}||})^2 (C_j^s)^{-1}$$

and set $Y_j^s = \sum_i f_i Z_{ij}^s$. The gradient of H_{trln2} is

$$\nabla H_{trln2}(f, \alpha) = \{f_i \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} \frac{tr(ln(Y_j^s) Z_{ij}^s)}{tr(Z_{ij}^s)}\}_{i=1}^{k}.$$

The optimization problem (6) is solved by gradient descent algorithm, which shows relatively fast convergence, unfortunately not always to the global minimum, because $H_{trln2}(f, \alpha)$ is not convex in $f \in P_k^+$.

Log-likelihood function gives us another choice for H,

$$H_{lik}(f; \alpha) = \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} \{tr(\Sigma[f]_j (C_j^s)^{-1}) - ln|\Sigma[f]_j (C_j^s)^{-1}| - n\} =$$

$$\sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} \{tr(Y_j^s) - ln|Y_j^s| - n\}.$$

The gradient of H_{lik} is

$$\nabla H_{lik}(f; \alpha) = \{f_i \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} \frac{tr((Y_j^s - I_n) Z_{ij}^s)}{tr(Z_{ij}^s)}\}_{i=1}^{k}.$$

Note that h_{lik} is neither symmetric nor satisfies the triangular inequality, but its importance is determined by the relation to normal distributions and its analytical properties. Define the matrix

$$B = \{b_{ij} = (d(q_j, p_i) - \frac{\pi}{2})^2\}_{i=1,j=1}^{k,k}.$$

Proposition 3 *If B has full rank, rank(B) = k, then for all* α, $H_{lik}(f; \alpha)$ *is a convex function in* P_k^+.

Proof: We have

$$\frac{\partial H_{lik}}{\partial f_i} = \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} tr(Z_{ij}^s - Z_{ij}^s (Y_j^s)^{-1}).$$

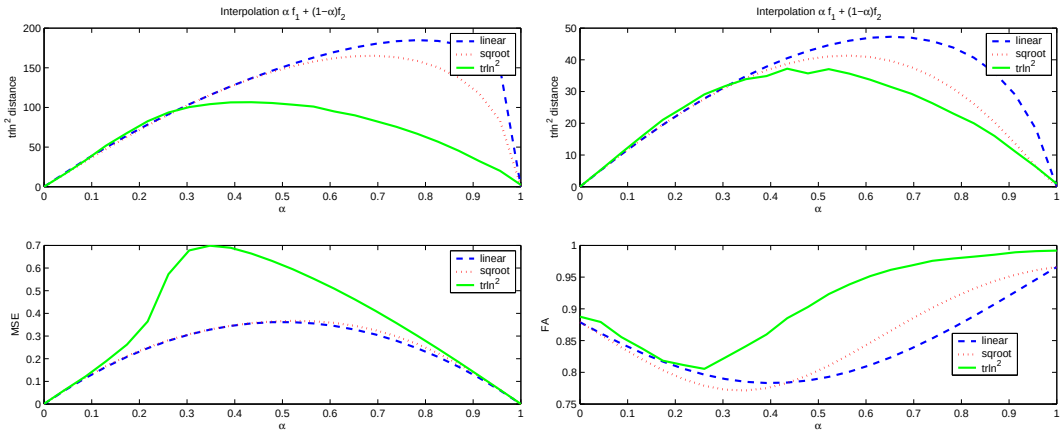

Figure 1: Two examples of interpolation of *pmfs* on $\mathbb{S}^2$ using h_{trln2}. The linear and square root interpolations are also given for reference. Top plots show H_{trln2} and H_{lik} for the three interpolations. Bottom plots show corresponding MSEs in the left and FAs in the right.

and

$$\frac{\partial^2 H_{lik}}{\partial f_i \partial f_l} = \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} tr(Z_{ij}^s (Y_j^s)^{-1} Z_{lj}^s (Y_j^s)^{-1}).$$

We want to show that the matrix of second partial derivatives is positive definite. Let $w = \{w_i\} \in \mathbb{R}^k$ and $w \neq 0$, then

$$\sum_{i,l} w_i w_l \frac{\partial^2 H_{lik}}{\partial f_i \partial f_l} = \sum_{s=1}^{m} \alpha_s \sum_{j=1}^{k} tr(\sum_{i=1}^{k} w_i Z_{ij}^s (Y_j^s)^{-1})^2 > 0,$$

since by the assumption for B, for at least one j, $\sum_{i=1}^{k} w_i Z_{ij}^s \neq 0$. □

The rank of B can be calculated using the pairwise distances between q and p points and only in very special circumstances this rank will be less than k. More formally, if a random process chooses the points, then $P(rank(B) < k) = 0$.

4. Examples and conclusions

Figure 1 shows interpolation between two $pmfs$ of size 6 ($m = 2, k = 6$) applying h_{trln2}. We compare it to the linear and the square root interpolations. It is also informative to compare the Mean-Squared Error (MSE) between different interpolations. It is defined by $MSE(\hat{f}) = \sum_{s=1}^{2} \alpha_s \sum_{i=1}^{k} (\hat{f}_i - f_i^s)^2$. Linear and square root interpolations, by their nature, are very close in MSE, but very different from $\hat{f}_{trln^2}(\alpha)$, which manifests the non-linear origin of the latter.

Another performance criteria relevant to the study of spherical data is the Fractional Anisotropy (FA). Let $\{\lambda_i\}_{i=1}^n$ be the eigenvalues of $\sum_{i=1}^{k} \overrightarrow{p_i}\overrightarrow{p_i}' f_i$, where $\overrightarrow{p_i}$ are considered vectors in $\mathbb{R}^n$ (thus FA is defined only for distributions on $\mathbb{S}^{n-1}$). Then we define $FA(f) = \{\frac{n}{n-1} \sum_{i=1}^{n} (\lambda_i - \bar{\lambda})^2 / \sum_{i=1}^{n} \lambda_i^2\}^{1/2}$. Fractional Anisotropy measures a distribution concentration. The higher FA, the more concentrated is the distribution about particular axes. A uniform distribution has $FA = 0$. As we may expect the linear interpolation substantially reduces the FA index. h_{trln2}-based one however, is more conservative and manage to sustain higher FA. Preserving the concentration factor is of importance for processing ODFs in HARDI, and the empirical evidence for the good FA performance of h_{trln2} is encouraging.

A second set of examples in figure 2 illustrates interpolation based on the likelihood function, h_{lik}. As we showed, this choice guarantees the convexity of H_{lik} and thus the continuity of the optimal solution $\hat{f}_{lik}(\alpha)$.

The likelihood based interpolation $\hat{f}_{lik}$ exhibits behaviour similar to that of $\hat{f}_{trln2}$. Again, it is very distinct from the linear and square-root ones and tends to preserve the anisotropy.

In conclusion, the proposed interpolation approach is general enough to be applied for distributions on any Riemannian manifold. Moreover, by employing a great variety of instruments, the similarity invariants, our method

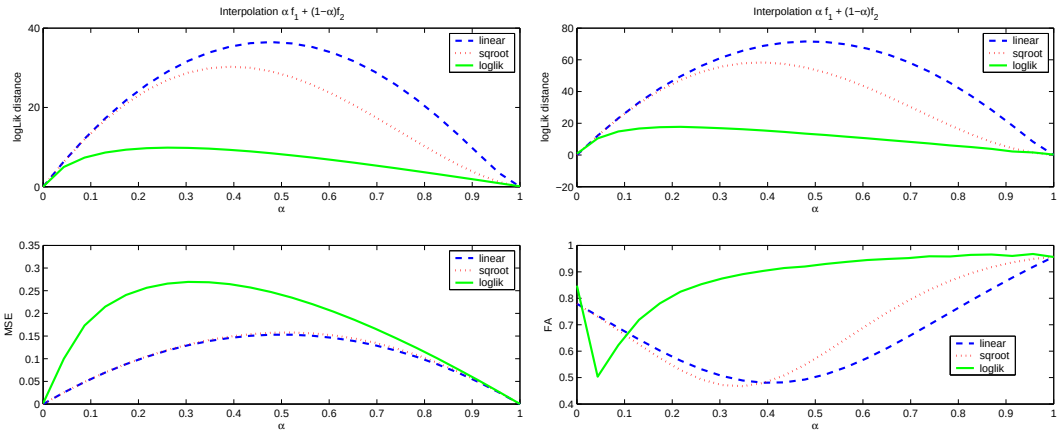

Figure 2: Two examples of interpolation of *pmfs* on $\mathbb{S}^2$ using h_{lik}. The linear and square root interpolations are also given for reference. Top plots show H_{lik} for the three alternatives. Bottom plots show corresponding MSEs in the left and FAs in the right.

allows application specific choices and provides flexibility.

5. Acknowledgements

We thank Vic Patrangenaru for the helpful comments and useful discussions. We also thank Paul Thompson from UCLA School of Medicine for providing us with a HARDI volume for research purposes.

References

[1] Arsigny,V., Fillard, P., Pennec X., and Ayache, N. (2007) Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328-347.

[2] Balov, N. (2008) Covariance fields. arXiv:0807.4690v1.

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

[4] Batchelor, P., Moakher, M., Atkinson, D., Calamante, F., Connelly, A. (2005) A rigorous framework for diffusion tensor calculus. Magn. Reson. Med., 53(1), 221-225.

[5] Chiang MC, Barysheva M, Lee AD, Madsen SK, Klunder AD, Toga AW, McMahon KL, de Zubicaray GI, Meredith M, Wright MJ, Srivastava A, Balov N, Thompson PM . Mapping Genetic Influences on Brain Fiber Architecture with High Angular Resolution Diffusion Imaging (HARDI), ISBI 2008.

[6] Dryden, I., Koloydenko, A., and Zhou, D., (2008) Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. University of Nottingham, NG7 2RD, UK.

[7] Fletcher, P. T., Joshi, S., (2007) Riemannian geometry for the statistical analysis of diffusion tensor data. Signal Processing, 87(2):250-262.

[8] Kent, J. (1982) The Fisher-Bingham distribution on the sphere. J. Royal Stat Soc. 44, 7180.

[9] Krieger L, N. C., Juul J., D., Conradsen, K. (1994) On the statistical analysis of orientation data. Acta Cryst., A50, 741-748

[10] Mardia, K. Directional statistics and shape analysis. Research Report STAT95/24, University of Leeds.

[11] Ohara, A., Suda, N., Amari., S. (1996) Dualistic Differential Geometry of positive definite matrices and its applications to related problems. *Linear Algebra and Its Applications* 247:31:53.

[12] Pennec. X., Fillard, P., Ayache, N (2006), A Riemannain framework for tensor computing. *Int. J. Computational Vision,* 66(1):41-66.

[13] Srivastava, A., Jermyn, I., Joshi, S. (2007) Riemannian Analysis of Probability Density Functions with Applications in Vision *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*