An efficient recursive estimator of the Fréchet mean on a hypersphere with applications to Medical Image Analysis

Hesamoddin Salehian[1], Rudrasis Chakraborty[1], Edward Ofori[2], David Vaillancourt[2], and Baba C. Vemuri[1] *

[1] Department of CISE, University of Florida, Gainesville, Florida, USA **
{salehian, rudrasis, vemuri}@cise.ufl.edu
[2] Department of Applied Physiology and Kinesiology, University of Florida, Florida, USA
{eofori,vcourt}@ufl.edu

Abstract. Finding the Riemannian center of mass or the Fréchet mean (FM) of manifold-valued data sets is a commonly encountered problem in a variety of fields of Science and Engineering including but not limited to, Medical Image Computing, Machine Learning, and Computer Vision. For instance, it is encountered in tasks such as, atlas construction, clustering, principal geodesic analysis etc. Traditionally, algorithms for computing the FM of the manifold-valued data require that the entire data pool be available apriori and not incrementally. When encountered with new data, the FM needs to be recomputed over the entire pool, which can be computationally as well as storage inefficient. A computational and storage efficient alternative is to consider a recursive algorithm for computing the FM which simply updates the previously computed FM when presented with a new data set. In this paper, we present such an alternative called the incremental Fréchet mean estimator (*iFME*) for data on the hypersphere. We prove the asymptotic convergence of *iFME* to the true FM of the underlying distribution from which the data samples were drawn. Further, we present several experiments demonstrating the performance on synthetic and real data sets.

1 Introduction

With the advent of sophisticated sensing technologies, manifold-valued data sets have become pervasive in many fields of applied sciences and Engineering including Medical Image Computing, Machine Learning and Computer Vision. Among these data, the most widely encountered are those that lie on a $k-$sphere, $k \geq 2$. To mention a few, the directional data which are often encountered in Image Processing and Computer Vision are points on the unit $2-$sphere $\mathbb{S}^2$ [15]. Further, 3×3 rotation matrices can be parameterized by unit quaternions which can be represented by points on the 3-dimensional unit sphere $\mathbb{S}^3$ [9]. Also, any probability density function, e.g., Orientation Distribution Function (ODF) in diffusion magnetic resonance imaging (MRI) [23], can be represented as points on a unit Hilbert sphere [4,20].

---

* Corresponding author.
** This research was supported in part by the NIH grant NS066340 to BCV

In most of these applications, mean computation is a key ingredient. Examples include, the interpolation and smoothing of ODF fields [5,4,8], estimation of the mean rotation from several corresponding pairs of points in multi-view geometry [9] and statistical analysis of directional data [15]. Given a set of samples on $\mathbb{S}^k$, the *Fréchet mean* (FM), is defined as the minimizer of the sum of squared geodesic distances. In general, the minimizer is non-unique and this issue has been well studied in literature and we refer the reader to [1,16] and references therein for details. It is also known that for a set of more than two samples on a hypersphere, the FM cannot in general be computed in closed form, and iterative schemes like the gradient descent must be employed [1,16] which for very large data sets can prove to be computationally quite expensive. Further, in many real-world applications the entire input data are not available all at once, and the population is usually augmented over time. Hence, in this context the standard gradient descent based iterative computation of the FM suffers from two major drawbacks: (1) for each new sample, it has to compute the new FM from scratch, and (2) it requires the entire input data to be stored, in order to estimate the new FM. Instead, an *incremental* i.e., a recursive technique can address this problem more efficiently with respect to time/space utility.

Recently, several incremental mean estimators for manifold-valued data have been reported. In [21], Sturm presented an incremental mean, the so called *inductive mean*, and proved its convergence to the true FM for all non-positively curved (NPC) spaces. In [7], authors showed several algorithms (including a recursive algorithm) for FM computation for data residing in CAT(0) spaces, which are NPC. They also demonstrated several applications of the same to Computer Vision and Medical Imaging. Further, in [10] an incremental FM computation algorithm along with its convergence and applications was presented for a population of Symmetric Positive Definite (SPD) matrices. Recently, in [14], Lim presented an inductive FM to estimate the weighted FM of SPD matrices. The convergence analysis in all of these works is applicable only to the samples belonging to NPC spaces and hence, their convergence analysis does not apply to the case of the hypersphere which is a positively curved Riemannian manifold with constant sectional curvature [11]. In [3], Arnaudon et al. present a stochastic gradient descent algorithm for barycenter computation of probability measures on Riemannian manifolds under some conditions. They also proved that their algorithm almost surely converges to the true Riemannian barycenter. Their algorithm is a stochastic version of ours as well as that of Sturm [21].

In this paper, we present a novel incremental FM estimator (*iFME*) of a set of samples on the hypersphere. When encountered with a new sample data set, an incremental update of the previously estimated FM is more computationally efficient compared to the non-incremental counterpart (*henceforth denoted by nFM*), because the update problem involves just the weighted FM of two items (previously computed mean and the new sample) and no optimization method is needed for its computation. This leads to significant efficiency in time and space (storage) consumption. Further, we will analytically show that in the limit (over the number of samples), our estimator converges to the true FM of the distribution from which the samples are drawn. To the best of our knowledge, this is the first convergence analysis for an incremental FM estimator on a positively curved Riemannian manifold. Finally, we will present examples of recursive

FM computation on several synthetic and real data sets along with its application to an incremental principal geodesic analysis *iPGA* algorithm which is used in the classification of movement disorder patients from their diffusion MR scans.

2   Riemannian Geometry of the Hypersphere

The hypersphere is the simplest of the constant positive curvature Riemannian manifolds encountered in numerous application problems. Its geometry is well known and here we will simply present the closed form expressions for the Riemannian Exponential and Log maps as well as the geodesic between two points on it. Further, we also present the square root parametrization of probability density functions, which allows one to identify them with points on the unit Hilbert sphere. This will be needed in representing the probability density functions namely, the ensemble average propagators (EAPs) in diffusion MRI, as points on the unit Hilbert sphere.

Without loss of generality we restrict the analysis to PDFs defined on the interval $[0, T]$ for simplicity: $\mathcal{P} = \{p : [0, T] \to \mathbb{R} | \forall s, p(s) \geq 0, \int_0^T p(s)ds = 1\}$. In [17], the Fisher-Rao metric was introduced to study the Riemannian structure of a statistical manifold (the manifod of probability densities). For a PDF $p_i \in \mathcal{P}$, the Fisher-Rao metric is defined as $\langle v_j, v_k \rangle = \int_0^T v_j(s)v_k(s)p_i(s)ds$, where $v_j, v_k \in T_{p_i}\mathcal{P}$. The Fisher-Rao metric is invariant to reparameterizations of the functions. In order to facilitate easy computations when using Riemannian operations, the square root density representation $\psi = \sqrt{p}$ was used in [20]. The space of square root density functions is defined as $\Psi = \{\psi : [0, T] \to \mathbb{R} | \forall s, \psi(s) \geq 0, \int_0^T \psi^2(s)ds = 1\}$. As we can see, Ψ forms a convex subset of the unit sphere in a Hilbert space. Then, the Fisher-Rao metric can be written as $\langle v_j, v_k \rangle = \int_0^T v_j(s)v_k(s)ds$ where, $v_j, v_k \in T_{\psi_i}\Psi$ are tangent vectors. Given any two functions $\psi_i, \psi_j \in \Psi$, the geodesic distance between these two points is given in closed form by $d(\psi_i, \psi_j) = \cos^{-1}(\langle \psi_i, \psi_j \rangle)$ The geodesic at ψ_i with a direction $v \in T_{\psi_i}\Psi$ is defined as $\gamma(t) = \cos(t)\psi_i + \sin(t)\frac{v}{|v|}$. Then, the Riemannian exponential map can be expressed as $exp_{\psi_i}(v) = \cos(|v|)\psi_i + \sin(|v|)\frac{v}{|v|}$, where, $|v| \in [0, \pi)$. The Riemannian logarithmic map is then given by $log_{\psi_i}(\psi_j) = u \cos^{-1}(\langle \psi_i, \psi_j \rangle)/\sqrt{\langle u, u \rangle}$ where, $u = \psi_j - \langle \psi_i, \psi_j \rangle \psi_i$.

Using the geodesic distance provided above, one can define the Fréchet mean (FM) of a set of points on the hypersphere as the minimizer of the sum of squared geodesic distances (so called Fréchet functional). Let $B(C, \rho)$, be the geodesic ball centered at C with radius ρ, i.e., $B(C, \rho) = \{Q \in \mathbb{S}^k | d(C, Q) < \rho\}$. Authors in [13] showed that for any $C \in \mathbb{S}^k$ and for data samples in $B(C, \frac{\pi}{2})$, the minimizer of the Fréchet functional exists and is unique. Therefore, in the rest of the paper, we assume that this condition is satisfied for any set of given points, $X_i \in \mathbb{S}^k$. For more details on Riemannian geometry of the sphere, reader is referred to chapter 2 of [11] and references therein.

3   Weak Consistency of *iFME* on the Sphere

In this section, we present the detailed proof of convergence of our recursive estimator on $\mathbb{S}^k$. The proposed method is similar in "spirit" to the incremental arithmetic mean

update in the Euclidean space; given the old mean, M_{n-1}, and the new sample, X_n, we define the new mean, M_n, as the weighted mean of M_{n-1} and X_n with the weights being $\frac{n-1}{n}$ and $\frac{1}{n}$, respectively. From a geometric viewpoint, this corresponds to the choice of the point on geodesic curve between M_{n-1} and X_n, with the parameter $t = \frac{1}{n}$.

Formally, let $X_1, X_2, ..., X_N$ be a set of N samples on hypersphere $\mathbb{S}^k$, all of which belong to the geodesic ball of radius $\frac{\pi}{2}$). The *iFME* estimate M_n of the FM with the nth given sample X_n is defined by:

where $A \#_t B$ is the point on the short-est geodesic path from A to B ($\in \mathbb{S}^k$) for a parameter value of t, and $\frac{1}{n}$ is the weight assigned to the new sample point (in this case the nth sample), which is

$$M_1 = X_1 \qquad (1)$$
$$M_n = M_{n-1} \#_{\frac{1}{n}} X_n \qquad (2)$$

henceforth called the *Euclidean weight*. In the rest of this section, we will show that if the number of given samples, N, tends to infinity, the *iFME* estimates will converge to the FM of the distribution from which the samples are drawn. *Note that the proof steps given below are not needed to compute the iFME, these steps are needed only to prove the weak consistency of iFME.* Our proof is based on the idea of projecting the samples on the sphere, X_i, to the tangent plane using the Gnomonic Projection [9], and perform the convergence analysis on the projected samples in this linear space, i.e., $\mathbf{x_i}$, instead of doing the analysis on the hypersphere. We take advantage of the fact that the geodesic curve between any pair of points on the hemisphere, is projected to a straight line in the tangent space at the anchor point (in this case, without loss of generality, assumed to be the north pole), via the gnomonic projection. A figure depicting the Gnomonic Projection is shown in Fig. 1.

Despite the simplifications used in the statis-tical analysis of the *iFME* estimates on the hy-persphere using the gnomonic projection, there is one important obstacle that must be considered. Without loss of generality, suppose the true FM of the input samples, X_i, is the north pole. Then, it can be shown through counter examples that:

Fig. 1: Gnomonic Projection

- The use of Euclidean weights, $\frac{1}{n}$, to update the *iFME* estimates on $\mathbb{S}^k$, does **not** necessarily correspond to the same weighting scheme between the old arithmetic mean and the new sample, in the projection space i.e., the tangent space.

The above fact can be illustrated us-ing two sample points on a unit circle ($\mathbb{S}^1$), $X_1 = \pi/6$ and $X_2 = \pi/3$, whose intrinsic mean is $M = \pi/4$. Then, the midpoint of the gnomonic projections of X_1 and X_2, which are denoted by $\mathbf{x_1}$ and $\mathbf{x_2}$, is $\hat{\mathbf{m}} = \frac{tan(\pi/3)+tan(\pi/6)}{2} = 1.1547 \neq tan(\pi/4) = \mathbf{m}$ (see Fig. 2).

For the rest of this section, without loss of generality, we assume that the true FM of N given samples is located at the
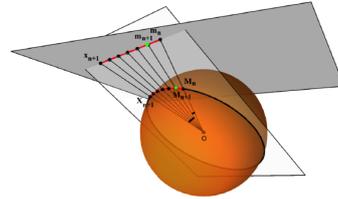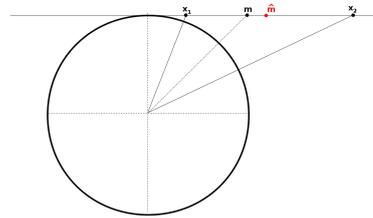

Fig. 2: Illustration of the counterexample

north pole. Since the gnomonic projection space is anchored at the north pole, this assumption leads to significant simplifications in our convergence analysis. However, a similar convergence proof can be developed for any arbitrary location of the FM, with the tangent (projection) space anchored at the location of this mean.

In what follows, we prove that the use of Euclidean weights, i.e., $w_n = \frac{1}{n}$, to update the incremental FM on the hypersphere, corresponds to a set of weights in the projection space, denoted henceforth by t_n, for which the weighted incremental mean in the tangent plane, converges to the true FM on the hypersphere, which in this case is the point of tangency.

Theorem 1 (Angle Bisector Theorem). *[2] Let M_n and M_{n+1} denote the iFME estimates for n and $n + 1$ given samples, respectively, and X_{n+1} denotes the $(n + 1)^{st}$ sample. Further, let $\mathbf{m_n}, \mathbf{m_{n+1}}, \mathbf{x_{n+1}}$ be the corresponding points in the projection space, then*

$$t_n = \frac{||\mathbf{m_n} - \mathbf{m_{n+1}}||}{||\mathbf{x_{n+1}} - \mathbf{m_{n+1}}||} = \frac{||O - \mathbf{m_n}||}{||O - \mathbf{x_{n+1}}||} \times \frac{\sin(d(M_n, M_{n+1}))}{\sin(d(M_{n+1}, X_{n+1}))} \tag{3}$$

where, $d(.)$ is the geodesic distance on the hypersphere.

In the rest of this section, we assume that the input samples, X_i, are within the geodesic ball, $B(C, \phi)$, where $0 < \phi < \pi/2$. This is needed for the uniqueness of the FM on the hypersphere (see [1]). Then, we bound t_n with respect to the radius ϕ.

Lemma 1 (Lower and Upper Bounds for t_n). *With the same assumptions made as in Theorem 1, the following inequality holds:*

$$\frac{\cos(\phi)}{n} \leq t_n \leq \frac{1}{\cos(\phi)^3 n} \tag{4}$$

Lower Bound. To prove this lower bound for t_n, we find the lower bounds for each fraction on the right hand side of Eq. 3. The first term reaches its minimum value, if M_n is located at the north pole, and X_{n+1} is located on the boundary of the geodesic ball, $B(C, \phi)$. In this case, $||O - \mathbf{m_n}|| = 1$ and $||O - \mathbf{x_{n+1}}|| = \frac{1}{\cos(\phi)}$. This implies that:

$$\frac{||O - \mathbf{m_n}||}{||O - \mathbf{x_{n+1}}||} \geq \cos(\phi) \tag{5}$$

Next, note that based on the definition of *iFME*, this second fraction in 3 can be rewritten as $\frac{\sin(d(M_n, M_{n+1}))}{\sin(d(M_{n+1}, X_{n+1}))} = \frac{\sin(d(M_n, M_{n+1}))}{\sin(n \times d(M_n, M_{n+1}))} = \frac{1}{U_{n-1}(\cos(d(M_n, M_{n+1})))}$ where, $U_{n-1}(x)$ is the Chebyshev polynomial of the second kind [19]. For any $x \in [-1, 1]$, the maximum of $U_{n-1}(x)$ is reached when $x = 1$, for which $U_{n-1}(1) = n$. Therefore, $U_{n-1}(x) \leq n$ and $\frac{1}{U_{n-1}(x)} \geq \frac{1}{n}$. This implies that:

$$\frac{\sin(d(M_n, M_{n+1}))}{\sin(n \times d(M_{n+1}, M_{n+1}))} = \frac{1}{U_{n-1}(\cos(d(M_n, M_{n+1})))}$$
$$\geq \frac{1}{n} \tag{6}$$

The inequalities 5 and 6, complete the proof.   ∎

Note that when ϕ tends to zero, $\cos(\phi)$ converges to one, and this lower bound tends to $\frac{1}{n}$, which is the case in Euclidean space.

Upper Bound. First, the upper bound for the first term in 3 is reached when M_n is on the boundary of geodesic ball, and X_{n+1} is given at the north pole. Therefore,

$$\frac{||O - \mathbf{m_n}||}{||O - \mathbf{x_{n+1}}||} \leq \frac{1}{\cos(\phi)} \tag{7}$$

Finding the upper bound for the $\sin$ term however is quite involved. Note that the maximum of the angle between OM_n and OX_{n+1}, denoted by α, is reached when M_n and X_{n+1} are both on the boundary of the geodesic ball, i.e., $\alpha \leq 2\phi$. Therefore, $\phi \in [0, \frac{\pi}{2})$ implies that $\alpha \in [0, \pi)$. Further, we show in the Appendix that the following inequality holds for any $\alpha \in (0, \pi)$.

$$\frac{\sin(\frac{n\alpha}{n+1})}{\sin(\frac{\alpha}{n+1})} \geq n\cos^2(\frac{\alpha}{2}) = n\cos^2(\phi) \tag{8}$$

From 7 and 6, the result follows. ∎

Thus far, we have shown analytical bounds for the sequence of weights, t_n, in the projection space, corresponding to Euclidean weights on sphere (Eq. 4). We now prove the convergence of *iFME* estimates to the true FM of distribution from which the samples are drawn, when the number of samples tend to infinity.

Theorem 2 (Unbiasedness). *Let (σ, ω) denote a probability space with probability measure ω. A vector valued random variable, $\mathbf{x}$ is a measurable function on σ taking values in $\mathbb{R}^k$, i.e., $\mathbf{x} : \sigma \to \mathbb{R}^k$. The distribution of $\mathbf{x}$ is the push-forward probability measure, $dP(\mathbf{x}) = \mathbf{x}^*(\sigma)$ on $\mathbb{R}^k$. The expectation is defined by $E[\mathbf{x}] = \int_\sigma \mathbf{x} d\omega$. Let $\mathbf{x_1}, \mathbf{x_2}, ...$ be i.i.d. samples from the distribution of $\mathbf{x}$. Also, let $\mathbf{m_n}$ be the incremental mean estimate corresponding to n^{th} given sample, $\mathbf{x_n}$, which is defined by: (i) $\mathbf{m_1} = \mathbf{x_1}$, (ii) $\mathbf{m_n} = t_n\mathbf{x_n} + (1 - t_n)\mathbf{m_{n-1}}$. Then, $\mathbf{m_n}$ is an unbiased estimator of the expectation $E[\mathbf{x}]$.*

Proof. For $n = 2$; $\mathbf{m_2} = t_2\mathbf{x_2} + (1-t_2)\mathbf{x_1}$, hence $E[\mathbf{m_2}] = t_2E[\mathbf{x}] + (1-t_2)E[\mathbf{x}] = E[\mathbf{x}]$.

By induction hypothesis we have, $E[\mathbf{m_{n-1}}] = E[\mathbf{x}]$. Then, $E[\mathbf{m_n}] = t_nE[\mathbf{x}] + (1 - t_n)E[\mathbf{x}] = E[\mathbf{x}]$, hence the result. ∎

Theorem 3 (Weak Consistency). *Let $var[\mathbf{m_n}]$ denotes the variance of the nth incremental mean estimate (which is defined in Theorem 2), with $\frac{\cos(\phi)}{n} \leq t_n \leq \frac{1}{\cos(\phi)^3 n}, \forall \phi \in [0, \pi/2)$. Then, $\exists p \in (0, 1]$, such that $\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} \leq (n^p\cos^6(\phi))^{-1}$.*

First note that $var[\mathbf{m_n}] = t_n^2var[\mathbf{x}] + (1 - t_n)^2var[\mathbf{m_{n-1}}]$. Since, $0 \leq t_n \leq 1$, one can see that $var[\mathbf{m_n}] \leq var[\mathbf{x}]$ for all n. Besides, for each n, the maximum of the right hand side is achieved, when t_n attains either its minimum or its maximum value. Therefore, we need to prove the theorem for the following two values of t_n, (i) $t_n = \frac{\cos(\phi)}{n}$ and (ii) $t_n = \frac{1}{n\cos^3(\phi)}$. These two cases will be proved in the Lemmas 2 and 3 respectively.

Lemma 2. *Suppose the same assumptions as in Theorem 2 are made. Further,* $t_n = \frac{1}{n\cos^3(\phi)}$, $\forall n$ *and* $\forall \phi \in [0, \pi/2)$, *then* $\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} \leq (n\cos^6(\phi))^{-1}$.

Proof. For $n = 1$, $var[\mathbf{m_1}] = var[\mathbf{x}]$ which yields the result, since $\cos(\phi) \leq 1$. Now, assume by induction that $\frac{var[\mathbf{m_{n-1}}]}{var[\mathbf{x}]} \leq (n-1)\cos^6(\phi))^{-1}$. Then,

$$
\begin{aligned}
\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} &= t_n^2 + (1-t_n)^2 \frac{var[\mathbf{m_{n-1}}]}{var[\mathbf{x}]} \leq t_n^2 + (1-t_n)^2 \frac{1}{(n-1)\cos^6(\phi)} \\
&\leq \frac{1}{\cos^6(\phi)n^2} + (1 - \frac{1}{\cos^3(\phi)n})^2 \times \frac{1}{(n-1)\cos^6(\phi)} \\
&\leq \frac{1}{\cos^6(\phi)n^2} + (1 - \frac{1}{n})^2 \times \frac{1}{(n-1)\cos^6(\phi)} \\
&= \frac{1}{\cos^6(\phi)n^2} + \frac{n-1}{n^2\cos^6(\phi)} = \frac{1}{n\cos^6(\phi)}
\end{aligned}
\tag{9}
$$

∎

Lemma 3. *Suppose the same assumptions as in Theorem 2 hold. Further,* $t_n = \frac{\cos(\phi)}{n}$, $\forall n$ *and* $\forall \phi \in [0, \pi/2)$, *then,* $\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} \leq n^{-p}$ *for some* $0 < p \leq 1$..

Proof. For $n = 1$, $var[\mathbf{m_n}] = var[\mathbf{x}]$ which yields the result, since $\cos(\phi) \leq 1$. Now, assume by induction that $\frac{var[\mathbf{m_{n-1}}]}{var[\mathbf{x}]} \leq (n-1)^{-p}$. Then,

$$
\begin{aligned}
\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} &= t_n^2 + (1-t_n)^2 \frac{var[\mathbf{m_{n-1}}]}{var[\mathbf{x}]} \leq t_n^2 + (1-t_n)^2 \frac{1}{(n-1)^p} \\
&\leq \frac{\cos^2(\phi)}{n^2} + \frac{(n-\cos(\phi))^2}{n^2} \times \frac{1}{(n-1)^p} \\
&= \frac{(n-1)^p\cos^2(\phi) + \cos^2(\phi) - 2n\cos(\phi) + n^2}{n^2(n-1)^p}
\end{aligned}
\tag{10}
$$

Now, it suffices to show that the numerator of the above expression is not greater than $n^{2-p}(n-1)^p$. In other words:

$$
(n-1)^p\cos^2(\phi) + \cos^2(\phi) - 2n\cos(\phi) + n^2 - n^{2-p}(n-1)^p \leq 0
\tag{11}
$$

The above quadratic function in $\cos(\phi)$ is less than zero, when

$$
n\left(\frac{1 - (n-1)^{p/2}\sqrt{(\frac{n-1}{n})^p + \frac{1}{n^p} - 1}}{1 + (n-1)^p}\right) \leq \cos(\phi) \leq n\left(\frac{1 + (n-1)^{p/2}\sqrt{(\frac{n-1}{n})^p + \frac{1}{n^p} - 1}}{1 + (n-1)^p}\right)
\tag{12}
$$

The inequality on the right is satisfied for all values of the cos function. Besides, it is easy to see that the function on the left hand side is increasing w.r.t. $n > 1$, and hence attains its minimum when $n = 2$. This implies that:

$$
\begin{aligned}
1 - \sqrt{2^{1-p} - 1} &\leq \cos(\phi) \\
\rightarrow \phi &\leq \cos^{-1}(1 - \sqrt{2^{1-p} - 1}) \\
\rightarrow 0 < p &\leq 1 - log_2[(1 - \cos(\phi))^2 + 1]
\end{aligned}
\tag{13}
$$

Note that $p > 0$, for all $\phi < \pi/2$.   ∎

Convergence. Armed with the above two results, it is easy to see that $\forall \phi \in [0, \pi/2)$, there exists a p satisfying $0 < p \leq 1$, such that

- If $t_n = \frac{\cos(\phi)}{n}$, then $\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} \leq \frac{1}{n^p} \leq \frac{1}{n^p \cos^6(\phi)}$, because $\cos(\phi) \leq 1$.
- If $t_n = \frac{1}{n \cos^3(\phi)}$, then $\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} \leq \frac{1}{n \cos^6(\phi)} \leq \frac{1}{n^p \cos^6(\phi)}$, because $p \leq 1$.

These two pieces together complete the proof of convergence.   ∎

The inequality in Theorem 3 implies that when $n \to \infty$, for any $\phi \in [0, \pi/2)$ the variance of *iFME* estimates in the projection space tends to zero. Besides, when ϕ approaches $\pi/2$, the corresponding power of n, as well as $\cos(\phi)$, become very small, hence the rate of convergence gets slower. Note that instead of the weights scheme used here (i.e., in spirit of incremental mean in Euclidean space), one can choose different weights scheme inherent to the manifold (i.e., as a function of curvature) to speed up the convergence rate.

4   Experimental Results

We now evaluate the effectiveness of the *iFME* algorithm, compared to the non-incremental counterpart, *nFM*, for computing the FM of a finite set of samples on the sphere (northern hemi-sphere not including the equator). As mentioned earlier, *nFM* for computing the FM uses a gradient descent technique to minimize the sum of squared geodesic distances cost function. We report the results for samples drawn from a mixture of Log-Normal distribution on the upper hemi-sphere. A set of random samples are drawn from the distribution and fed to both the *iFME* and the *nFM* algorithms, incrementally. The computation time needed by each method for computing the sample FM, and the error was recorded, for each new sample incrementally introduced. The error is defined by the geodesic distance between the estimated mean (using either *iFME* or the *nFM*) and the true expected value of the input distribution. Because of the randomness in generating the samples, we repeated this experiment 100 times for each case, and the mean time consumption and the error for each method are shown. All the computation time required for various algorithms reported in this paper, were measured on an Intel-7 quad-core processor, 25GB RAM desktop.

A set of samples are drawn from a mixture of Log-Normal distributions on the sphere. The mean of each Log-Normal component is set randomly, and the covariance matrices are set to $0.1I$ and $0.2I$ respectively, where I is the identity matrix. Similar to the previous experiment, the
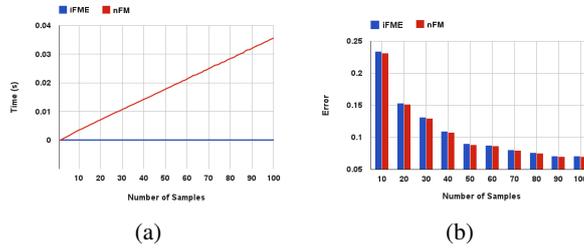

Fig. 3: Time and error comparisons of *iFME* and *nFM*

performances of *iFME* and *nFM* are evaluated with respect to the time consumption and accuracy, and are illustrated in Fig. 3a and Fig. 3b respectively.

Though *iFME*'s accuracy is still very similar to that of *nFM*, it estimates the intrinsic mean significantly faster. Now, we compare the performance of *iFME* with *nFM* and *eFME* (the extrinsic FM estimator). The *eFME* is defined as $\Pi(\sum_i X_i)$, where Π is the projection operator from $\mathbb{R}^{k+1}$ to $\mathbb{S}^k$, i.e., $\Pi(\mathbf{x}) = \mathbf{x}/|\mathbf{x}|$. We randomly generated 50 samples on northern hemi-sphere with varying data
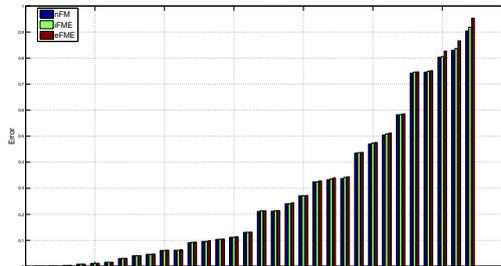

Fig. 4: Comparison between *iFME*, *nFM*, *eFME*

variance. From Fig. 4, it is evident that for high data variance, *eFME* exhibits high computationally efficiency but a rather poor accuracy in its estimate of the FM.

5 Application to the classification of movement disorders

In this section we first present a novel incremental version of the PGA algorithm in [25] applicable to data lying on a sphere. We will call this the *iPGA* algorithm. Then, we present an application of *iPGA* to real data sets. The (batch-mode) PGA proposed in [25] for diffusion tensor fields consists of (1) computing the FM of the input data, (2) projecting each data point to the tangent space at the FM using the Riemannian log-map, (3) performing standard PCA in the tangent plane and (4) projecting the result (principal vectors) back to the manifold, using the Riemannian exp-map.

An incremental form of this PGA technique was proposed recently in [18]. However, their technique was limited to manifolds with non-positive sectional curvatures. Equipped with the *iFME* on the sphere presented in the previous section, we can now extend the *iPGA* technique in [18] to the case when data lie on a hypersphere. For this, we need to use *iFME* for the FM computation and use the parallel transport operation on the hypersphere. The parallel transport operation on the hypersphere can be expressed in a closed form expression. The formula for parallel transporting $\mathbf{p} \in T_{\mathbf{n}}\mathbb{S}^k$ from $\mathbf{n}$ to $\mathbf{m}$ is given by

$$\mathbf{q} = \Gamma_{\mathbf{n}\to\mathbf{m}}(\mathbf{p})$$
$$= \left(\mathbf{p} - \mathbf{v}\left(\frac{\mathbf{v}^t\mathbf{p}}{||\mathbf{v}||^2} \right) \right) + \frac{\mathbf{v}^t\mathbf{p}}{||\mathbf{v}||^2} \left(\mathbf{n}\left(-\sin(||\mathbf{v}||)||\mathbf{v}|| \right) + \mathbf{v}\cos(||\mathbf{v}||) \right)$$

where, $\mathbf{v} = Log_{\mathbf{n}}\mathbf{m}$. We now present the *iPGA* method in an algorithm form summarized in Table 1 and refer the reader for details to [18].

The dataset for classification contains HARDI scans from (1) healthy controls, and patients with, (2) Parkinson's disease (PD), and (3) essential tremor (ET). We aim to automatically discriminate between these three classes using features derived from the HARDI data. This dataset consists of 25 controls, 24 PD, and 15 ET images. The

HARDI data were acquired using a 3T Phillips MR scanner with the following parameters: $TR = 7748\,ms$, $TE = 86\,ms$, $b-$values: 0, 1000 $\frac{s}{mm^2}$, 64 gradient directions and voxel size = $2 \times 2 \times 2\,mm^3$.

Authors in [24] employed DTI based analysis, using scalar-valued features to address the problem of movement disorder classification. Later in [18], a PGA-based classification algorithm was proposed, using Cauchy deformation tensors (computed from a non-rigid registration of patient scans to a HARDI atlas) which are SPD matrices. In the next subsection, we develop classification method based on (1)

1: **Input** the data matrix $\mathbb{A}_k = [\mathbf{v}_1, ..., \mathbf{v}_k]$ for k samples
the new sample $\mathbf{x}_{k+1}$, and the old mean $\mathbf{m}_k$
2: Compute $\mathbf{m}_{k+1}$ from $\mathbf{x}_{k+1}$ and $\mathbf{m}_k$, using Eq. 2
3: $\mathbf{y}_{k+1} = Log_{\mathbf{m}_{k+1}}(\mathbf{x}_{k+1})$
4: Parallel Transport $\mathbf{z}_{k+1} = \Gamma_{\mathbf{m}_{k+1} \to \mathbf{n}}(\mathbf{y}_{k+1})$
5: Compute $\mathbf{r}_{k+1} = Log_{\mathbf{m}_{k+1}}(\mathbf{m}_k)$ and $\mathbf{t}_{k+1} = \Gamma_{\mathbf{m}_{k+1} \to \mathbf{n}}(\mathbf{r}_{k+1})$
6: Add $\mathbf{t}_{k+1}$ to every column of $\mathbb{A}_k$ to obtain $\hat{\mathbb{A}}_k = [\hat{\mathbf{v}}_1, ..., \hat{\mathbf{v}}_\mathbf{k}]$
7: Perform standard PCA on $\mathbb{A}_{k+1} = \left[\hat{\mathbb{A}}_k \vdots \mathbf{z}_{k+1} \right]$
8: Parallel transport the jth principal component, $\mathbf{p}_j$,
back to $T_{\mathbf{m}_{k+1}} \mathbb{S}^\mathbf{k}$, via $\mathbf{q}_j = \Gamma_{\mathbf{n} \to \mathbf{m}_{k+1}}(\mathbf{p}_j)$

Table 1: The Incremental PGA Algorithm on a Unit Hypersphere

Ensemble Average Propagators (EAPs) derived from HARDI data within an ROI, and (2) shapes of the ROI over the input population. Using a square root density parameterization [22], both features can be mapped to points on an unit Hilbert sphere, where the proposed *iFME* in conjunction with the *iPGA* method is applicable.

Classification Results using the Ensemble Average Propagator as Features: To capture the full diffusional information, we chose to use the ensemble average propagator (EAP) at each voxel as our feature in the classification. We compute the EAPs using the method described in [12] and use the square root density parameterization of each EAP. This way the full diffusion information at each voxel is represented as a point on the unit Hilbert sphere.

We now present the classification algorithm which is a combination of *iPGA*-based reduced representation and the nearest neighbor classifier. The input to the *iPGA* algorithm is EAP features in this case. The input HARDI data are first rigidly aligned to the atlas computed from the normal group, then a 3-D box surrounding the ROI, i.e., the midbrain, is placed on each image, and the EAPs within this box are computed. Finally, the EAP field extracted from each ROI image is identified with a point on the product manifold (the number of elements in the product is equal to the number of voxels in the ROI) of unit Hilbert spheres. This is in spirit similar to the case of the product manifold formalism in [25,18].

A set of 10 Control, 10 PD and 5 ET images are randomly picked as the test set, and the rest of the images are used for training. Also, classification is performed using *iPGA*, PGA and the standard PCA, and is repeated 300 times to report the average accuracy. The results using EAP features are summarized in Table 2. It is evident that the accuracy of *iPGA* is roughly the same as that of the non-incremental PGA, while both methods are considerably more accurate than the standard PCA, as they account for the non-linear geometry of the sphere. Further, the savings in computation time for *iPGA* are significant in comparison to PGA as evident from the table.

	Results using Shape Features									Results using EAP Features								
	Control vs. PD			Control vs. ET			PD vs. ET			Control vs. PD			Control vs. ET			PD vs. ET		
	iPGA	PGA	PCA	iPGA	PGA	PCA	iPGA	PGA	PCA	iPGA	PGA	PCA	iPGA	PGA	PCA	iPGA	PGA	PCA
Accuracy	91.5	93.0	67.3	88.3	90.1	75.7	86.1	87.6	64.6	92.7	93.5	59.8	90.1	91.3	70.2	89.7	90.9	66.0
Sensitivity	88.0	91.0	52.0	84.4	86.2	80.1	80.5	82.4	58.4	90.7	91.8	48.3	87.5	89.7	79.8	84.0	84.7	56.3
Specificity	95.0	95.0	82.7	92.2	94.1	71.3	91.7	92.8	70.8	94.7	95.2	71.3	92.7	92.9	60.6	95.5	97.1	75.7
Time (s)	4.1	18.5		4.0	14.2		3.5	14.7		11.6	30.9		9.8	27.3		10.8	28.0	

Table 2: Classification results from *iPGA*, PGA and PCA respectively.

Classification Results using the Shapes as Features: In this section, we evaluated the *iPGA* algorithm based on shape of the Substantia Nigra region in the brain images, for the task of movement disorder classification. We first collected random samples (point) on the boundary of each 3-D shape, and applied the Schrodinger distance transform (SDT) technique in [6] to represent each shape as a point on the unit hyper-sphere. The size of the ROI for the 3-D shape of interest was set to $28 \times 28 \times 15$, the resulting samples lie on a $\mathbb{S}^{11759}$ manifold. Then, we used *iPGA* for classification. The results given in Table 2 show significant time gains for *iPGA* over PGA but with similar accuracy.

6   Conclusion

In this paper, we presented a novel incremental Fréchet mean estimator (*iFME*), for data lying on a hypersphere. We proved the asymptotic convergence of *iFME* to the true FM. Significant time efficiency of *iFME* compared to *nFM* was shown via synthetic and real data experiments. Further, we also presented an incremental PGA (*iPGA*) algorithm that entailed the use of *iFME*. We used the *iPGA* in conjunction with a nearest neighbor classifier to classify movement disorder patients using diffusion MR brain scans. Our classification demonstrated significant gains in computation time compared to batch mode *PGA* (in conjunction with the nearest neighbor classifier), and as expected achieved the same accuracy as batch-mode PGA. In our future work, we will focus on providing an upper bound on the distance between iFME and the FM for finite set of samples.

Appendix

In this appendix, we show that $\frac{\sin(\frac{n\alpha}{n+1})}{\sin(\frac{\alpha}{n+1})} \geq n\cos^2(\frac{\alpha}{2})$ for any $\alpha \in (0, \pi)$.

Proof. Let, $f = sin(n\theta) - ncos^2(\frac{n+1}{2}\theta)sin(\theta)$, $\theta \in (0, \alpha/(n+1))$, $\alpha \in (0, \pi)$, $n \geq 1$. and $f_\theta = n\,cos(n\theta) + 2n\,cos(\frac{n+1}{2}\theta)\,sin(\theta)\,sin(\frac{n+1}{2}\theta)\,(\frac{n+1}{2}) - n\,cos^2(\frac{n+1}{2}\theta)\,cos(\theta)$ Solving this eqn., as $\theta \in (0, \pi/(n+1))$, we get, $\theta = 0$. But, $f_{\theta\theta}|_{\theta=0} = 0$. Hence, we check $f_{\theta\theta\theta}$.

$f_{\theta\theta\theta}|_{\theta=0} = -n^3 + 1.5n\,(n+1)^2 + n \quad > 0, \quad n \geq 1$ So, at $\theta = 0$, f has a minimum where, $\theta \in (0, \alpha/(n+1))$. $f|_{\theta=0} = 0$ Thus, $f \geq 0$ as $n \geq 1$. As for $\theta \in (0, \alpha/(n+1))$, $sin(\theta) > 0$, $\frac{f}{sin(\theta)} \geq 0$ $\frac{f}{sin(\theta)} = \frac{sin(n\theta)}{sin(\theta)} - ncos^2(\frac{n+1}{2}\theta)$ Hence, $\frac{sin(n\theta)}{sin(\theta)} - ncos^2(\frac{n+1}{2}\theta) \geq 0$ Then, by substituting $\theta = \alpha/(n+1)$, we get $\frac{\sin(\frac{n\alpha}{n+1})}{\sin(\frac{\alpha}{n+1})} \geq n\cos^2(\frac{\alpha}{2})$

∎

References

1. B. Afsari, R. Tron, , et al. On the convergence of gradient descent for finding the riemannian center of mass. *SIAM J. on Control and Optimization*, 51(3):2230–2260, 2013.
2. G. Amarasinghe. On the standard lengths of angle bisectors and the angle bisector theorem. *Global Journal of Advanced Research on Classical and Modern Geometries*, 1(1), 2012.
3. M. Arnaudon, C. Dombry, A. Phan, and L. Yang. Stochastic algorithms for computing means of probability measures. *Stochastic Processes and their Applications*, 122(4):1437–1455, 2012.
4. H. E. Cetingul, B. Afsari, et al. Group action induced averaging for HARDI processing. In *ISBI*, pages 1389–1392, 2012.
5. J. Cheng, A. Ghosh, et al. A riemannian framework for orientation distribution function computing. In *MICCAI*, pages 911–918. Springer, 2009.
6. Y. Deng, A. Rangarajan, et al. A riemannian framework for matching point clouds represented by the schrodinger distance transform. In *CVPR*, 2014.
7. A. Feragen, S. Hauberg, M. Nielsen, and F. Lauze. Means in spaces of tree-like shapes. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 736–746. IEEE, 2011.
8. A. Goh, C. Lenglet, et al. A nonparametric riemannian framework for processing high angular resolution diffusion images and its applications to odf-based morphometry. *NeuroImage*, pages 1181–1201, 2011.
9. R. Hartley, J. Trumpf, et al. Rotation averaging. *IJCV*, 103(3):267–305, 2013.
10. J. Ho, G. Cheng, et al. Recursive karcher expectation estimators and geometric law of large numbers. In *AISTATS*, pages 325–332, 2013.
11. B. Iversen. *Hyperbolic geometry*, volume 25. Cambridge University Press, 1992.
12. B. Jian and B. C. Vemuri. A unified computational framework for deconvolution to reconstruct multiple fibers from DWMRI. *IEEE TMI*, 26:1464–1471, 2007.
13. W. S. Kendall. Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proceedings of the London Mathematical Society*, 3(2):371–406, 1990.
14. Y. Lim and M. Pálfia. Weighted inductive means. *LAA*, 453:59–83, 2014.
15. K. V. Mardia and P. E. Jupp. *Directional statistics*, volume 494. John Wiley & Sons, 2009.
16. X. Pennec. Probabilities and statistics on riemannian manifolds: Basic tools for geometric measurements. In *NSIP*, pages 194–198, 1999.
17. C. R. Rao. Differential metrics in probability spaces. *Differential geometry in statistical inference*, 10:217–240, 1987.
18. H. Salehian, D. Vaillancourt, et al. iPGA: Incremental principal geodesic analysis with applications to movement disorder classification. In *MICCAI*, pages 765–772. 2014.
19. N. J. Sloane et al. The on-line encyclopedia of integer sequences, 2003.
20. A. Srivastava, I. Jermyn, and S. Joshi. Riemannian analysis of probability density functions with applications in vision. In *CVPR*, pages 1–8, 2007.
21. K. T. Sturm. Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, 2003.
22. J. Sun, Y. Xie, et al. Dictionary learning on the manifold of square root densities and application to reconstruction of diffusion propagator fields. In *IPMI*, pages 619–631, 2013.
23. D. S. Tuch, T. G. Reese, et al. Diffusion mri of complex neural architecture. *Neuron*, 40(5):885–895, 2003.
24. D. Vaillancourt, M. Spraker, et al. High-resolution diffusion tensor imaging in the substantia nigra of de novo parkinson disease. *Neurology*, pages 1378–1384, 2009.
25. Y. Xie, B. C. Vemuri, and J. Ho. Statistical analysis of tensor fields. In *MICCAI*, pages 682–689. Springer, 2010.