Recursive Computation of the Fréchet Mean on Non-Positively Curved Riemannian Manifolds with Applications

Guang Cheng Jeffrey Ho Hesamoddin Salehian Baba C. Vemuri

Abstract Computing the Riemannian center of mass or the finite sample Fréchet mean has attracted enormous attention lately due to the easy availability of data that are manifold-valued. Manifold-valued data are encountered in numerous domains including but not limited to Medical Image Computing, Computer Vision, Machine Learning etc. It is common practice to estimate the finite sample Fréchet mean by using a gradient descent technique to find the minimum of the Fréchet function when it exists. The convergence rate of this gradient descent method depends on many factors including the step size and the variance of the given manifold-valued data etc. As an alternative to the gradient descent technique, we propose a recursive (incremental) algorithm for estimating the Fréchet mean/expectation (iFEE) of the distribution from which the sample data are drawn. The proposed algorithm can be regarded as a geometric generalization of the well known incremental algorithm for computing arithmetic mean, since it reinterprets this algebraic formula in terms of geometric operations on geodesics in the more general manifold setting. In particular, given known formulas for geodesics, iFEE does not require any optimization in contrast to the non-incremental counterparts and offers significant improvement in efficiency and flexibility. For the case of simply connected, complete and non-positively curved Riemannian manifolds, we prove that iFEE converges to the true expectation in the limit. We present several experiments demonstrating the efficiency and accuracy of iFEE in comparison to the non-incremental counterpart

Guang Cheng,
Google, e-mail: seagle.99@gmail.com

Jeffrey Ho
MediaTek, e-mail: jho.jeffrey@gmail.com

Hesamoddin Salehian
University of Florida, e-mail: salehian@cise.ufl.edu

Baba C. Vemuri (corresponding author)
University of Florida e-mail: vemuri@cise.ufl.edu

for, computing the Fréchet mean of symmetric positive definite matrices, k-means clustering and diffusion tensor image segmentation tasks respectively.

Key words: Incremental algorithm, Fréchet expectation, law of large numbers, DTI Segmentation

1 Introduction

In computer vision and machine learning, statistical analysis of features often requires them to be considered as random variables. Although features are always represented as collections of real numbers, their idiosyncratic origins and indigenous constraints often can best be interpreted not as vectorial features in $\mathbb{R}^n$ but as manifold features, features belonging to some embedded submanifold $\mathscr{M}$ of $\mathbb{R}^n$. In this sense, manifold-valued random variables abound in vision and machine learning literature: Popular image features such as SIFT and HOG, due to normalization, are often features defined on spheres. For applications involving directional and geometric data, important features can often be found in Lie groups that model their underlying symmetries (e.g., [1]). From Lie groups, one is naturally led to their homogeneous and symmetric spaces, and not surprisingly, the homogeneous and symmetric spaces of classical Lie groups such as (complex) projective spaces, Stiefel manifolds (of which spheres form an important family) and Grassmannians are especially rich domains for generating useful features, making their appearances in a wide range of vision problems, including shape and procrustean analysis [2], subspace clustering [3], dynamic texture classification [4], object recognition [5,6] and many others. In particular, as a symmetric space of the general linear group $\mathbf{GL}(n)$ (n-by-n nonsingular matrices), the manifold $\mathbb{P}(n)$ of n-by-n symmetric positive-definite matrices has featured prominently in many vision and learning problems, from the relatively simple image structure tensors [7] that has been the traditional workhorse in vision algorithms to the more refined covariance features used in tracking [8,9] and recognition [10], and from the application domain of Diffusion Tensor MRI (DT-MRI) in medical imaging (e.g., [11–14]) to the more esoteric domain of information geometry using Fisher-Rao metric [15]. While manifolds identify the domains on which the features are defined, probability and metric then furnish the tools for analyzing and characterizing their uncertainty and similarity respectively. In the general manifold setting, the union of probability and geometry (metric) naturally leads to the notion of Fréchet mean/expectation [16] of a random variable, whose definition requires the specifications of both a distribution and a Riemannian metric. Consequently, an important computational problem is to estimate the Fréchet mean using samples of the distribution. However, in this era of massive and continuous streaming data, samples are often given either as a whole that are difficult to utilize due to their size or in parts with availability depending on other external factors. Therefore, from an application viewpoint, the desired algorithm should be *incremental* in nature in order to maximize computational efficiency and account

for data availability, requirements that are seldom addressed in more theoretically oriented domains. In this chapter, we propose iFEE, a novel incremental algorithm for computing the Fréchet mean of manifold-valued random variables, and establish its convergence for simply connected, complete Riemannian manifolds with non-positive sectional curvature.

Let $\mathscr{M}$ denote a (complete) Riemannian manifold and $d\omega$ its canonical Riemannian volume measure. We will consider a probability measure dP on $\mathscr{M}$ that is absolutely continuous with respect to $d\omega$, i.e., its density function $P(\mathbf{x})$ exists with $dP = P(\mathbf{x})\,d\omega$ as the measure. The Fréchet expectation (or Fréchet mean) of the probability measure dP is defined as

$$\mathbf{m} = \arg\min_{\mathbf{z}\in\mathscr{M}} \int_{\mathscr{M}} \mathbf{d}^2_{\mathscr{M}}(\mathbf{z},\mathbf{x})\,P(\mathbf{x})d\omega, \tag{1}$$

where $\mathbf{d}_{\mathscr{M}}(\mathbf{z},\mathbf{x})$ is the Riemannian geodesic distance between $\mathbf{z},\mathbf{x} \in \mathscr{M}$. Note that the expectation $\mathbf{m}$ is a point in $\mathscr{M}$ (or a set in general), and it requires the specifications of the Riemannian metric $\mathbf{d}^2_{\mathscr{M}}(\mathbf{z},\mathbf{x})$ and the probability measure dP. In general, the existence and uniqueness of Fréchet expectation for an arbitrary probability measure defined on a Riemannian manifold is a subtle and technical issue; however, for simply connected and complete Riemannian manifolds with non-positive sectional curvature, a theorem of E. Cartan (see section 6.1.5 in [17]) shows that the Riemannian center of mass (Fréchet expectation) always exists and is unique for any probability measure absolutely continuous with respect to ω. Therefore, in this chapter, we will focus exclusively on such Riemannian manifolds and in particular, we will define an estimator $\mathbf{m}_k$ of $\mathbf{m}$ for any sequence $\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_k$ of i.i.d. samples drawn from the distribution dP and show that $\mathbf{m}_k$ converges asymptotically to $\mathbf{m}$ as $k \to \infty$.

In the Euclidean domain, the estimator $\mathbf{m}_k$ is well-known and it is simply the average of the (finite) samples

$$\mathbf{m}_k = \frac{\mathbf{x}_1 + ... + \mathbf{x}_k}{k}. \tag{2}$$

The validity of the estimator of course is guaranteed by the (weak) law of large numbers, which states that the estimator $\mathbf{m}_k$ converges in probability to the true mean $\mathbf{m}$. For practitioners in computer vision and machine learning (and others), this well-known result is so deeply ingrained that many times we use it without immediate awareness of it. In particular, $\mathbf{m}_k$ can be computed incrementally using the formula

$$\mathbf{m}_{k+1} = \frac{k\,\mathbf{m}_k + \mathbf{x}_{k+1}}{k+1}, \tag{3}$$

and in $\mathbb{R}^n$, the two formulas above are in fact equivalent. However, due to their algebraic appearances, the underlying geometry of the two formulas are often overlooked, and on non-Euclidean manifolds where the relations between geometry and algebra are no longer as transparent, proper generalizations of these two formulas must rely on their geometric interpretations rather than their algebraic forms.

Specifically, Equation 2 can be generalized geometrically to any Riemannian manifold $\mathscr{M}$ as the center-of-mass of the finite samples, according to

$$\mathbf{m}_k = \arg\min_{\mathbf{z}\in\mathscr{M}} \sum_{i=1}^{k} \mathbf{d}^2_{\mathscr{M}}(\mathbf{z}, \mathbf{x}_i), \tag{4}$$

and computationally, it can be solved as an optimization problem on $\mathscr{M}$ for each k. On the other hand, the incremental form in Equation 3 involves only two points and in a Euclidean space $\mathbb{R}^n$, we can interpret it geometrically as moving an appropriate distance away from $\mathbf{m}_k$ towards $\mathbf{x}_{k+1}$ on the straight line joining $\mathbf{x}_{k+1}$ and $\mathbf{m}_k$. This geometric procedure can be readily extended to any Riemannian manifold using geodesics, and for classical spaces such as the aforementioned examples, there are often closed-form formulas for geodesics joining two given points. This readily yields an algorithm for computing $\mathbf{m}_k$ that does not require any function optimization, a considerable advantage often realized as gains in computation time of several orders in magnitude over non-incremental algorithms based on Equation 4. However, because the presence of curvature, generalizations of Equations 2 and 3 are no longer equivalent as in the Euclidean case, and the incremental computation of $\mathbf{m}_k$ will also depend on the ordering of the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_k$, with the latter non-commutative property marking the qualitative divergence between the Euclidean and non-Euclidean cases. In particular, it is not immediately clear that $\mathbf{m}_k$ will indeed converge asymptotically to the true expectation $\mathbf{m}$.

In this chapter, we establish the convergence of the incremental algorithm for simply connected complete Riemannian manifolds with non-positive sectional curvature, and the proof presented in Section 2 is geometric in nature and elementary in detail. The basic idea is to use the Riemannian exponential and logarithm maps to transfer the problem from the manifold $\mathscr{M}$ to the tangent space $\mathbf{T}_{\mathbf{m}}$ at $\mathbf{m}$, and in the tangent space, we can compare the magnitudes of the incremental mean $\mathbf{m}_{k+1}$ computed by iFEE and the Euclidean mean computed by Equation 3. An important geometric consequence of the non-positive sectional curvature assumption is that the former is always not larger than the latter, and this provides the desired contraction that can be deployed in an inductive argument similar to the one for proving the law of large numbers in the Euclidean domain. In particular, the proof illustrates and illuminates the effect of curvature on the convergence of the incremental estimator, with the convergence in the Euclidean case (zero curvature) implicitly implying the convergence for all negative curvature cases. Furthermore, the convergence result also provides a geometric generalization of the law of large numbers in that the well-known sample average in the Euclidean law of large numbers is now replaced by the geometric operation of moves on geodesics.

In the statistics literature, manifold-valued random variables, or more general metric space-valued random variables have been studied quite extensively, e.g., [18–23]. However, their focus has always been on establishing and characterizing convergence of finite-sample means, and the attention is on extending Equation 2 to more general domains. However, in our more application-oriented context, the focus is instead on generalizing the incremental form in Equation 3, and this provides

a context that is qualitatively different from the aforementioned works in statistics. Perhaps the most well-known example is the incremental Principal Component Analysis (PCA) [24] and its various applications in Computer Vision and Image Processing (e.g., [25]) that demand and hence motivate the incrementalization of a well-known method originated from pure statistics. In particular, computation of the mean from sample data plays an important role in a variety of applications such as clustering, segmentation and atlas construction in medical imaging, and the performances of these algorithms are often determined by how efficiently and accurately can the mean be computed. As an incremental algorithm, iFEE provides a far more computationally efficient alternative to the standard non-incremental algorithms that compute the Fréchet mean based on minimizing Equation 4. The (asymptotic) accuracy and efficiency of iFEE have been thoroughly evaluated using experiments with synthetic and real data, and in the experiment section of this chapter, we report the significant gains in running time achieved by iFEE over other non-incremental methods without any noticeable degradation in its accuracy.

2 Algorithm and Convergence Analysis

In this section, we present the incremental algorithm for computing the Fréchet expectation on general Riemannian manifolds and provide a convergence proof of the incremental algorithm for simply connected and complete Riemannian manifolds with non-positive sectional curvature. Although phrased in the context of Riemannian geometry, the incremental algorithm and most of the convergence proof do not require much beyond the elementary notions such as Riemannian exponential and logarithm maps that are familiar to a large section of the Computer Vision and Medical Image Computing audience. For simplicity of exposition, we will assume $\mathscr{M}$ is a simply connected and complete Riemannian manifold with non-positive sectional curvature such that the Riemannian exponential map $\mathbf{Exp}_{\mathbf{x}}$ and its inverse, the Riemannian logarithm map $\mathbf{Log}_{\mathbf{x}}$, based at every $\mathbf{x} \in \mathscr{M}$ are diffeomorphisms between the tangent space $\mathbf{T}_{\mathbf{x}}$ and $\mathscr{M}$. In particular, $\mathscr{M}$ is assumed to have the $\mathbb{R}^n$ topology[1] and by Hopf-Rinow Theorem [26], any two points $\mathbf{x}, \mathbf{y}$ in $\mathscr{M}$ can be joined by a unique geodesic whose length gives the Riemannian distance $\mathbf{d}_{\mathscr{M}}(\mathbf{x}, \mathbf{y})$ between $\mathbf{x}$ and $\mathbf{y}$. An important example of a complete Riemannian manifold with non-positive sectional curvature topologically equivalent to $\mathbb{R}^k$ for some k is the space $\mathbb{P}(n)$ of n-by-n symmetric positive matrices equipped with the affine-invariant Riemannian metric.

Let $\mathbf{x}$ denote an $\mathscr{M}$-valued random variable and dP its associated probability measure (distribution) on $\mathscr{M}$ that is absolutely continuous with respect to the Riemannian volume measure $d\mathbf{x}$ with density function $P(\mathbf{x})$. Since $\mathbf{Exp}_{\mathbf{x}}$ is a diffeomorphism between $\mathbf{T}_{\mathbf{x}}$ and $\mathscr{M}$, we can use $\mathbf{Exp}_{\mathbf{x}}$ to pull the distribution dP (and its

[1] This topological assumption is not particularly restrictive since by Cartan-Hadamard Theorem [17], any simply connected complete d-dimensional Riemannian manifold with non-positive sectional curvature can be constructed topologically from $\mathbb{R}^d$.

density function) back to $\mathbf{T_x}$, essentially using $\mathbf{Exp_x}$ to identify distributions on $\mathscr{M}$ and $\mathbf{T_x}$. When there is no possibility of confusion, we will use the same notation for such a pair of distributions[2]. By a theorem of E. Cartan (see page 256 and proposition 60 on page 234 in [17]), dP has a unique (Fréchet) expectation $\mathbf{m}$ defined by Equation 1, and as the minimum, the first-order stationary condition gives

$$\int_{\mathscr{M}} \mathbf{Log_m}(\mathbf{x})\, P(\mathbf{x}) d\mathbf{x} = 0. \tag{5}$$

Using $\mathbf{Exp_m}$ and $\mathbf{Log_m}$ to identify distributions on $\mathscr{M}$ and $\mathbf{T_m}$, the integral above can be transferred onto $\mathbf{T_m}$

$$\int_{\mathbf{T_m}} x P(x) dx = 0, \tag{6}$$

where the distribution $P(x)dx$ on $\mathbf{T_m}$ in Equation 6 is the pull-back of the distribution $P(\mathbf{x})d\mathbf{x}$ on $\mathscr{M}$ in Equation 5 using $\mathbf{Exp_m}$.

2.1 The Incremental Fréchet Expectation Estimator (iFEE)

Let $\mathbf{x}_1, \mathbf{x}_2, ...$ be a sequence of i.i.d. samples of the probability distribution dP. The incremental Fréchet expectation estimator $\mathbf{m}_k, k = 1, 2, ...$ is defined as follows:

1. $\mathbf{m}_1 = \mathbf{x}_1$.
2. For $k > 1$, let $\gamma_k(t)$ denote the unique geodesic joining $\mathbf{m}_{k-1}$ and $\mathbf{x}_k$ such that $\gamma_k(0) = \mathbf{m}_{k-1}$ and $\gamma_k(1) = \mathbf{x}_k$; $\mathbf{m}_k = \gamma_k(\frac{1}{k})$.

Since the geodesic $\gamma_k(t)$ is constant speed [26], it follows from the definition of $\mathbf{m}_k$ that $\mathbf{d}_{\mathscr{M}}(\mathbf{x}_k, \mathbf{m}_k) = (k-1)\,\mathbf{d}_{\mathscr{M}}(\mathbf{m}_{k-1}, \mathbf{m}_k)$. In particular, when $k = 2$, $\mathbf{m}_2$ is simply the midpoint on the geodesic joining $\mathbf{x}_1$ and $\mathbf{x}_2$. For $k > 2$, the estimator $\mathbf{m}_k$ is closer to the previous estimator $\mathbf{m}_{k-1}$ than to the new sample $\mathbf{x}_k$ by a factor of $k-1$, a direct generalization of the Euclidean formula in Equation 3, interpreted geometrically. The incremental estimator $\mathbf{m}_k$ is also a $\mathscr{M}$-valued random variable, and $\mathbf{x}_i$ being independent samples implies that $\mathbf{m}_k$ is also independent of $\mathbf{x}_j$ for $j > k$. We will denote $dP(\mathbf{m}_k)$ its probability measure and $P(\mathbf{m}_k)\,d\mathbf{m}_k$ its density function with respect to the Riemannian volume measure ($\mathbf{m}_k$ in $d\mathbf{m}_k$ indicates the variable of integration).

The convergence result established below is a particular type of convergence called the convergence *in probability* of the random variables $\mathbf{m}_k$ to the expectation $\mathbf{m}$ (see [27]). Specifically, for this particular type of convergence, we need to show that for sufficiently large k, the probability that $\mathbf{d}^2_{\mathscr{M}}(\mathbf{m}_k, \mathbf{m})$ is larger than any $\varepsilon > 0$ can be made arbitrary small: for $\varepsilon, \delta > 0$, there exists an integer $K(\varepsilon, \delta)$ depending on ε, δ such that

$$\mathbf{Pr}\{\, \mathbf{d}^2_{\mathscr{M}}(\mathbf{m}_k, \mathbf{m}) > \varepsilon \,\} < \delta, \tag{7}$$

[2] Points in $\mathscr{M}$ are denoted by boldface font and their corresponding points in the tangent space are denoted by regular font.

for $k > K(\varepsilon, \delta)$. A simple application of the Markov's inequality then shows that it is sufficient to demonstrate that the mean squared error $\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k)$ of $\mathbf{m}_k$ with respect to $\mathbf{m}$ defined by

$$\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k) = \int_{\mathscr{M}} \mathbf{d}^2_{\mathscr{M}}(\mathbf{m}_k, \mathbf{m}) P(\mathbf{m}_k)\, d\mathbf{m}_k, \tag{8}$$

converges to zero as $k \to \infty$. Recall that for a nonnegative real-valued random variable $\mathbf{X}$, Markov's inequality states that

$$\mathbf{Pr}\{\mathbf{X} > \varepsilon\} \leq \frac{\mathbf{E}[\mathbf{X}]}{\varepsilon}, \tag{9}$$

where $\mathbf{E}[\mathbf{X}]$ is the expectation of $\mathbf{X}$. In our context, $\mathbf{X} = \mathbf{d}^2_{\mathscr{M}}(\mathbf{m}_k, \mathbf{m})$, and $\mathbf{E}[\mathbf{X}] = \mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k)$. Therefore, if $\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k) \to 0$ as $k \to \infty$, Equation 7 is true for any $\varepsilon, \delta > 0$ with $k > K(\varepsilon, \delta)$: since $\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k) \to 0$ as $k \to \infty$, there is a $K(\varepsilon, \delta)$ such that for all $k > K(\varepsilon, \delta)$, $\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k) = \mathbf{E}[\mathbf{X}] < \varepsilon\delta$, and Equation 7 follows readily from Equation 9.

The convergence proof will also rely on the variance of the random variable $\mathbf{x}$ defined by $\mathbf{Var}(\mathbf{x}) = \int_{\mathscr{M}} \mathbf{d}^2_{\mathscr{M}}(\mathbf{x}, \mathbf{m}) P(\mathbf{x})\, d\mathbf{x}$. In general, the integrals in $\mathbf{Var}(\mathbf{x})$ and $\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k)$ are difficult to evaluate directly on $\mathscr{M}$. However, considerable simplification is possible if the integrals are pulled back to the tangent space at $\mathbf{m}$. Specifically, since $\mathbf{d}^2_{\mathscr{M}}(\mathbf{m}_k, \mathbf{m})$ can be determined as the squared $\mathbf{L}^2$-norm of the tangent vector $\mathbf{Log}_{\mathbf{m}}(\mathbf{m}_k)$ in $\mathbf{T}_{\mathbf{m}}$, the integral above has a much simpler form on $\mathbf{T}_{\mathbf{m}}$

$$\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k) = \int_{\mathbf{T}_{\mathbf{m}}} \|m_k\|^2 P(m_k)\, dm_k. \tag{10}$$

We remind the reader that the measures $P(\mathbf{m}_k)\, d\mathbf{m}_k, P(m_k)\, dm_k$ in the two equations above, although defined on different domains, are identified via $\mathbf{Exp}_{\mathbf{m}}$.

2.2 The Example of $\mathbb{P}(n)$

From an application viewpoint, $\mathbb{P}(n)$ equipped with the affine (**GL**) invariant metric is certainly the most important example of complete Riemannian manifold of non-positive sectional curvature. Specifically, the general linear group $\mathbf{GL}(n)$ (n-by-n non-singular matrices) acts on $\mathbb{P}(n)$ according to the formula: $\forall \mathbf{g} \in \mathbf{GL}(n), \forall \mathbf{M} \in \mathbb{P}(n)$, $\mathbf{g}^*(\mathbf{M}) = \mathbf{g}\mathbf{M}\mathbf{g}^\top$. The action is not only transitive but for any pair of $\mathbf{M}, \mathbf{N} \in \mathbb{P}(n)$, there exists a $\mathbf{g} \in \mathbf{GL}(n)$ such that $\mathbf{g}\mathbf{M}\mathbf{g}^\top$ is the identity matrix and $\mathbf{g}\mathbf{N}\mathbf{g}^\top$ is a diagonal matrix, a geometric interpretation of the well-known algebraic fact that a pair of symmetric positive-definite matrices can be simultaneously diagonalized. The tangent space at each point in $\mathbb{P}(n)$ is identified with the vector space of n-by-n symmetric matrices, and a **GL**-invariant Riemannian metric [28] can be specified by the inner product for the tangent space $\mathbf{T}_{\mathbf{M}}$ of $\mathbf{M} \in \mathbb{P}(n)$: $< \mathbf{U}, \mathbf{V} >_{\mathbf{M}} = \mathbf{tr}(\mathbf{M}^{-1}\mathbf{U}\mathbf{M}^{-1}\mathbf{V})$, where $\mathbf{U}, \mathbf{V}$ are tangent vectors considered as sym-

metric matrices and **tr** denotes the trace of a matrix. The differential geometry of $\mathbb{P}(n)$ has been studied extensively by the differential geometer Helgason [28], and the invariance of the metric under **GL**-action makes many aspects of its geometry tractable. For example, $\mathbb{P}(n)$ is a complete Riemannian manifold with negative sectional curvature, and there is a closed-form formula for the Riemannian distance between $\mathbf{M},\mathbf{N} \in \mathbb{P}(n)$

$$\mathbf{d}^2_{\mathbb{P}(n)}(\mathbf{M},\mathbf{N}) = \mathbf{tr}(\mathbf{Log}(\mathbf{M}^{-1}\mathbf{N})^2) \tag{11}$$

where **Log** denotes the matrix logarithm. Furthermore, there is also a closed-form formula for the (unique) geodesic joining any pair of points $\mathbf{M},\mathbf{N}$ in $\mathbb{P}(n)$. Because of the invariance, the geodesic paths in $\mathbb{P}(n)$ originate from geodesic curves joining the identity matrix to diagonal matrices. Specifically, let $\mathbf{D}$ denote a diagonal matrix with positive diagonal entries $d_1, d_2, ..., d_n > 0$. The geodesic path $\gamma(t)$ joining the identity matrix and $\mathbf{D}$ is given by

$$\gamma(t) = \begin{pmatrix} e^{t\log(d_1)} & & 0 \\ & \ddots & \\ 0 & & e^{t\log(d_n)} \end{pmatrix} \quad or \quad \gamma(t) = \begin{pmatrix} d_1^t & & 0 \\ & \ddots & \\ 0 & & d_n^t \end{pmatrix}. \tag{12}$$

It is evident that $\gamma(0) = \mathbf{I}_n$ is the identity matrix and $\gamma(1) = \mathbf{D}$. In particular, $\gamma(t) = \mathbf{D}^t$, as the fractional power of the diagonal matrix. For a general symmetric positive-definite matrix $\mathbf{M}$, its fractional power can be defined knowing its eigen-decomposition: $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ with $\mathbf{D}$ diagonal and $\mathbf{U}$ orthogonal, its power $\mathbf{M}^t$ for $t \geq 0$ is $\mathbf{M}^t = \mathbf{U}\mathbf{D}^t\mathbf{U}^\top$. These matrix operations provide the steps for computing the geodesic joining any pairs of $\mathbf{M},\mathbf{N} \in \mathbb{P}(n)$: using $\mathbf{g} = \mathbf{M}^{-\frac{1}{2}}$, $(\mathbf{M},\mathbf{N})$ can be transformed simultaneously to $(\mathbf{I}_n, \mathbf{M}^{-\frac{1}{2}}\mathbf{N}\mathbf{M}^{-\frac{1}{2}})$, and the geodesic path $\gamma(t)$ joining $\mathbf{M},\mathbf{N}$ is the transform under $\mathbf{M}^{\frac{1}{2}}$ of the geodesic path $\overline{\gamma(t)}$ joining $(\mathbf{I}, \mathbf{M}^{-\frac{1}{2}}\mathbf{N}\mathbf{M}^{-\frac{1}{2}})$. Let $\mathbf{M}^{-\frac{1}{2}}\mathbf{N}\mathbf{M}^{-\frac{1}{2}} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ denote the eigen-decomposition of $\mathbf{M}^{-\frac{1}{2}}\mathbf{N}\mathbf{M}^{-\frac{1}{2}}$, and the geodesic path joining $(\mathbf{I}, \mathbf{M}^{-\frac{1}{2}}\mathbf{N}\mathbf{M}^{-\frac{1}{2}})$ is the transform under $\mathbf{U}$ of the geodesic path joining $\mathbf{I},\mathbf{D}$, with the latter being $\mathbf{D}^t$. Working backward, this gives the simple formulas $\overline{\gamma(t)} = (\mathbf{M}^{-\frac{1}{2}}\mathbf{N}\mathbf{M}^{-\frac{1}{2}})^t$, and $\gamma(t) = \mathbf{M}^{\frac{1}{2}}(\mathbf{M}^{-\frac{1}{2}}\mathbf{N}\mathbf{M}^{-\frac{1}{2}})^t\mathbf{M}^{\frac{1}{2}}$. Using the above formula, the iFEE has a particularly simple form: given the i.i.d. samples $\mathbf{x}_1, \mathbf{x}_2, ...$ in $\mathbb{P}(n)$, the incremental estimator is given by

$$\mathbf{m}_1 = \mathbf{x}_1 \tag{13}$$

$$\mathbf{m}_{k+1} = \mathbf{m}_k^{\frac{1}{2}}(\mathbf{m}_k^{-\frac{1}{2}}\mathbf{x}_{k+1}\mathbf{m}_k^{-\frac{1}{2}})^{\frac{1}{k+1}}\mathbf{m}_k^{\frac{1}{2}}. \tag{14}$$

We remark that the computation of $\mathbf{m}_{k+1}$ requires the eigen-decomposition of the symmetric positive-definite matrix $\mathbf{m}_k^{-\frac{1}{2}}\mathbf{x}_{k+1}\mathbf{m}_k^{-\frac{1}{2}}$ that can be done using many efficient and robust numerical algorithms.

2.3 Convergence Analysis

The analysis of the estimator's convergence rests on the asymptotic behavior of $\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k)$ defined in Equation 8 and its equivalent form defined on the tangent space $\mathbf{T}_{\mathbf{m}}$ in Equation 10. In particular, $\mathbf{m}_k$ *converges to* $\mathbf{m}$ *in the L2 norm* if $\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k) \to 0$ as $k \to \infty$ [27]. An important element in the proof of the latter is a crucial upper bound for the geodesic distance $\mathbf{d}_{\mathscr{M}}(\mathbf{m}_{k+1}, \mathbf{m})$ between $\mathbf{m}_{k+1}$ and $\mathbf{m}$ using the Euclidean distance in $\mathbf{T}_{\mathbf{m}}$, which is a consequence of the non-positive curvature assumption. The specific detail is illustrated in Figure 1 and given in the Proposition below. The basic idea is that by lifting $\mathbf{m}_k$ and $\mathbf{x}_{k+1}$ back to $\mathbf{T}_{\mathbf{m}}$ using the Riemannian logarithm map $\mathbf{Log}_{\mathbf{m}}$. A triangle Σ can be formed in $\mathbf{T}_{\mathbf{m}}$ using the three points m_k, x_{k+1} and $\mathbf{o}$, the origin. The geodesic distance $\mathbf{d}_{\mathscr{M}}(\mathbf{m}_{k+1}, \mathbf{m})$ is then bounded by the Euclidean distance between the origin $\mathbf{o}$ and the corresponding point $\overline{m}_k$ on Σ. This upper bound in terms of the Euclidean length of a vector in $\mathbf{T}_{\mathbf{m}}$ is useful because the resulting integral that gives the desired upper bound for the MSE in Equation 10 can be easily evaluated in an inductive argument using the distributions $P(m_{k-1})$ and $P(x)$ instead of $P(m_k)$. Specifically, we have

Proposition 1. *Let* $\mathbf{x}, \mathbf{y}, \mathbf{z}$ *be three points on a complete Riemannian manifold* $\mathscr{M}$ *with non-positive sectional curvature and* $\gamma(t)$ *the unique geodesic path joining* $\mathbf{x}, \mathbf{y}$ *such that* $\gamma(0) = \mathbf{x}, \gamma(1) = \mathbf{y}$. *Furthermore, let* $x = \mathbf{Log}_{\mathbf{z}}(\mathbf{x}), y = \mathbf{Log}_{\mathbf{z}}(\mathbf{y})$, *and* $\overline{\gamma}(t)$ *denote the straight line joining* x, y *such that* $\overline{\gamma(0)} = x, \overline{\gamma(1)} = y$. *Then,* $\mathbf{d}_{\mathscr{M}}(\gamma(t), \mathbf{z}) \leq \|\overline{\gamma(t)}\|$.

The proposition is a consequence of the non-positive assumption on curvature. Geometrically, it asserts that the geodesic distance between $\mathbf{m}$ and any point on the geodesic $\gamma(t)$ cannot be greater than the Euclidean distance in $\mathbf{T}_{\mathbf{m}}$ between the origin and the corresponding point on the line joining x and y. The proof of the proposition is relegated to the appendix.

Armed with the above proposition, the proof of the convergence of the iFEE is straightforward and essentially follows from the law of large numbers in the Euclidean case.

Theorem 1. *The incremental Fréchet expectation estimator* $\mathbf{m}_k$ *converges to the true Fréchet expectation* $\mathbf{m}$ *in probability, i.e., as* $k \to \infty$,

$$\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k) \to 0.$$

Proof. We will inductively show that

$$\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_k) \leq \frac{1}{k}\mathbf{Var}(\mathbf{x}). \tag{15}$$

Since $\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_1) = \mathbf{Var}(\mathbf{x})$, the inequality clearly holds for $k = 1$. For $k+1 > 1$, we have, by Proposition 1,

$$\mathbf{MSE}_{\mathbf{m}}(\mathbf{m}_{k+1}) \leq \int_{\mathbf{T}_{\mathbf{m}}} \|\frac{k m_k + x}{k+1}\|^2 P(m_k) P(x) dm_k dx.$$

The integral on the right can be evaluated as

$$\int_{\mathbf{T_m}} \|\frac{km_k+x}{k+1}\|^2 P(m_k)P(x)\,dm_k dx = \int_{\mathbf{T_m}} \frac{k^2}{(k+1)^2}\,\|m_k\|^2 P(m_k)\,dm_k$$
$$+2\frac{k}{(k+1)^2}\int_{\mathbf{T_m}}\int_{\mathbf{T_m}} m_k^\top x P(m_k)P(x)\,dm_k dx + \int_{\mathbf{T_m}} \frac{1}{(k+1)^2}\,\|x\|^2 P(x)\,dx$$
$$= \frac{k^2}{(k+1)^2}\mathbf{MSE_m}(\mathbf{m}_k) + \frac{1}{(k+1)^2}\mathbf{Var}(\mathbf{x}) \leq \frac{1}{k+1}\mathbf{Var}(\mathbf{x}),$$

where the last inequality follows from the induction hypothesis, and

$$\int_{\mathbf{T_m}}\int_{\mathbf{T_m}} m_k^\top x P(m_k)P(x)\,dm_k dx = 0$$

follows from Equation 6, although $\int_{\mathbf{T_m}} m_k P(m_k) dm_k$ is not guaranteed to be zero.

We remark that although Proposition-1 is valid for any other point $\mathbf{z} \neq \mathbf{m}$ in $\mathscr{M}$, the cross-term in the sum above $\int_{\mathbf{T_m}}\int_{\mathbf{T_m}} m_k^\top x P(m_k)P(x)\,dm_k dx$, when evaluated at $\mathbf{T_z}$ is generally nonzero. In particular, the above argument only works for the unique Fréchet expectation $\mathbf{m}$.

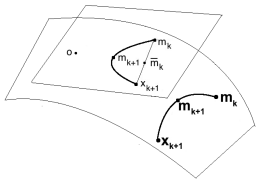


Fig. 1: Geometric consequences of non-positive curvature assumption. The three points $\mathbf{m}_k, \mathbf{x}_{k+1}, \mathbf{m}_{k+1}$ are on a geodesic $\gamma(t)$ in $\mathscr{M}$. After lifting back to the tangent space $\mathbf{T_m}$ at $\mathbf{m}$, the point $m_{k+1} = \mathbf{Log_m}(\mathbf{m}_{k+1})$ is closer to the origin than the corresponding point $\overline{m}_k$ on the straight line joining m_k and x_{k+1}. Note that the Euclidean distance between the origin and m_{k+1} is the same as the geodesic distance $\mathbf{d}_{\mathscr{M}}(\mathbf{m}_{k+1}, \mathbf{m})$.

3 Related Work

Manifold-valued random variables or more generally, random variables taking values on general metric spaces have been studied quite extensively in probability and statistics literature since the seminal work of Fréchet [16]. In the context of Riemannian manifolds, the notion of center of mass that is equivalent to the Fréchet expectation as defined by Equation 1 was initially introduced by E. Cartan (see section 6.1.5 in [17]), who established, among many other things, the uniqueness of Riemannian center of mass for complete manifolds of non-positive sectional curvature. For general Riemannian manifolds, the Fréchet expectation is unique only for distributions with some special properties, for example, when their supports are contained in convex geodesic balls. In statistics literature, the primary interest and focus are on establishing the convergence of finite-sample means to the true expec-

tation. With random variables taking values in general metric spaces and the non-uniqueness of expectation, characterizations of convergence require more elaborate machinery to account for the marked increase in the topological and geometric complexity. The basic form of the general law of large numbers for metric space-valued random variables was first established in [18, 19] under different assumptions on the metric spaces and the types of convergence. This abstract framework has been applied to study concrete statistical problems such as procrustean shape analysis [29], and several recent works [22, 23] have substantially extended the scope of these earlier works, both in abstraction and in application. In the context of Riemannian manifolds, the pair of papers [20, 21] provide some of the basic results, including a very general central limit theorem and several concrete examples concerning both the intrinsic and extrinsic means of several classical manifolds such as the complex projective spaces used in procrustean analysis. We remark that the idea of computing the mean incrementally and the question of its convergence do not seem to have been discussed nor studied in these works.

In the conference version of this chapter [30] we proved the convergence of the incremental estimator under the assumptions that the distribution is symmetric and the Riemannian manifold $\mathscr{M}$ is the space of symmetric positive-definite matrices $\mathbb{P}(n)$ equipped with the *GL*-invariant Riemannian metric. This result and its proof were later extended to general distributions on $\mathbb{P}(n)$ in our second conference paper [31]. The proofs presented in both papers rely on an inequality that is valid only for Riemannian manifolds with non-positive sectional curvature, and this makes the method ill-suited for further extension to manifolds with positive curvature that includes important examples in computer vision applications such as Grassmannians, Stiefel manifolds and most compact Lie groups. In particular, the connection between the Euclidean and non-Euclidean cases were not made explicit in these two earlier approaches, and the proof presented in this chapter that uses geometric comparisons is considerably more flexible in its potential for future extensions.

After the publication of [30] [31], we were made aware of the paper by Sturm [32] wherein he formulated and proved a substantially stronger convergence result for length spaces, a more general class of spaces than Riemannian manifolds. Specifically, length spaces are metric spaces that determine the distance between two points using the minimal length of a path joining them, and compared with Riemannian manifolds, length spaces retain the notion of geodesics (distance-minimizing paths) but forsake the manifold structure as well as the exponential and logarithm maps. Surprisingly, it is still possible, in the absence of a manifold structure, to define a useful notion of non-positive curvature for length spaces, and Strum [32] has formulated and proved a convergence result for length spaces of non-positive curvature, of which complete Riemannian manifolds of non-positive curvature are special cases. Although Sturm's result subsumes ours, the convergence theorem and its proof presented in [32] require considerably more machinery and longer exposition to compensate for the loss of familiar structures such as the Riemannian exponential and logarithm maps. From this perspective, our present exposition offers three important contributions: First, we present a significantly shorter and more accessible convergence proof using only elementary Riemannian geom-

etry that is familiar to general readership of the Computer Vision and Medical Image Computing communities. Second, our proof provides a more transparent and clear connection between the convergence in the Euclidean domain (the law of large numbers) and the non-Euclidean manifolds. Third, and most importantly, our proof method that relies heavily on the linearization provided by the Riemannian exponential and logarithm maps should be better adapted for studying the convergence problems for more general Riemannian manifolds and analyzing other related issues. Although the generality provided by [32] is both reassuring and welcoming, applications in Computer Vision and Machine Learning often require a specialized instead of generalized mathematical context, and in particular, general results must be sharpened and improved for special cases, perhaps in terms of shorter proofs or more precise characterizations, for which our current work serves as an example.

There are several articles in literature on computing the finite sample Fréchet mean on P_n. Moakher [33], Bhatia and Holbrook [34], presented methods for computing this intrinsic mean. Independently, there was work by Fletcher and Joshi [35] that presents a gradient descent algorithm for computing the finite sample Fréchet mean of diffusion tensors (matrices in P_n). Further, Ando, Li and Mathias (ALM) [36] presented a technique to compute the geometric mean of $n \geq 3$ SPD matrices and listed 10 properties (now called the ALM axioms) that their mean satisfied. It is to be noted that their mean is distinct from the Fréchet mean of the given sample set. In [37] Bini et al. present an extension of the unweighted geometric mean of two SPD matrices to higher number of SPD matrices by using "symmetrization methods" and induction, which satisfies the ALM axioms. This however is not the Fréchet mean of the given sample set. More recently, Lim and Pálfia [38] presented results on computing a weighted inductive mean of a finite set of SPD matrices. In [38], authors restrict themselves to discrete probability densities on P_n unlike the work presented in this chapter where we consider the continuous densities. For other types of means, we refer the reader to the excellent discussion in Pennec [39] wherein, a gradient descent algorithm is presented for finding the finite sample Fr'echet mean for simply connected non-positively curved Riemannian manifolds with curvature bounded from below. In [40], Afsari et al. present a comprehensive set of results on the convergence of the gradient descent algorithm for finding the Riemannian center of mass (a.k.a. finite sample Fréchet mean) on Riemannian manifolds.

In Medical Image Computing, an impetus for developing efficient algorithms for computing means from samples is provided by the prominent role played by the mean tensor (using various kinds of distances/divergences) in solving a wide range of important problems that include diffusion tensor image (DTI) as well as structure tensor field segmentation, interpolation, clustering and atlas construction. For example, authors in [41] generalize the geometric active contour based piece-wise constant segmentation [42, 43] to segmentation of DTIs using the Euclidean distance to measure the distance between two SPD tensors. Authors in [44] present a geometric active contour-based approach [45, 46] for tensor field segmentation that used information from the diffusion tensors to construct the so-called structure tensor which was a sum of structure tensors formed from each component of the

diffusion tensor. A Riemannian metric on the manifold of SPD matrices was used in [11, 47, 48] for DTI segmentation and for computing the mean interpolant of diffusion tensors respectively. In [14, 49, 50] and [47], the symmetrized KL-divergence (KL_s) was used for DTI segmentation and interpolation, respectively. Other applications of computing mean diffusion tensor field from a finite sample of diffusion tensor fields (structure tensor fields) can be found in [51, 52]. However, none of the above methods for computing the mean of SPD matrices (or SPD fields) used within the segmentation or the dictionary learning algorithms or in their own right for interpolation are in recursive form, even though it is evident that a recursive formulation would be more desirable in designing a more efficient algorithm.

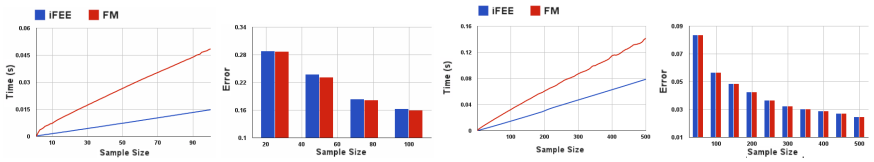
4 Experiments

In this section, we empirically demonstrate the accuracy and efficiency of the proposed iFEE algorithm using a set of experiments. The distributions studied in these experiments are all defined on $\mathbb{P}(n)$, the space of n-by-n symmetric positive-definite (SPD) matrices equipped with the *GL*-invariant Riemannian metric. All experiments reported in this section were performed on an Apple laptop with a 2.5GHz Intel Core *i*5 CPU and 4GB DDR3. All reported timings are on the aforementioned CPU.

4.1 Performance of the Incremental Fréchet Expectation Estimator

Symmetric Distributions: We illustrate the performance of iFEE on a set of random samples on $\mathbb{P}(n)$ drawn from a symmetric distribution, and compare the accuracy and computational efficiency of iFEE and the non-incremental (batch-mode) gradient descent algorithm for computing the finite sample Fréchet mean (FM) of the given dataset. To this end, a set of 100 i.i.d samples from a log-Normal distribution [53] on $\mathbb{P}(6)$ are generated, and the Fréchet mean is computed using iFEE as well as the batch-mode (gradient descent) method. We set the expectation and the variance of log-Normal distribution to the identity matrix and one, respectively. The error in estimation is measured by the geodesic distance from each estimated point to the identity. Further, for each new sample, the computation time for each method is recorded. Figure 2a illustrates the significant difference in running time between iFEE and the *batch mode method denoted using the legend, FM, in this figure and figures to follow*. It can be seen that the time taken by iFEE is considerably shorter than the batch mode (FM) method.

The accuracy errors of the two estimators are shown in Figure 2b. It can be seen that the incremental estimator provides roughly the same accuracy as the batch-mode counterpart. Furthermore, for large numbers of samples, the incremental estimation error clearly converges to zero. Therefore, the algorithm performs more accurately as the number of data samples grows.



(a) (b) (c) (d)

Fig. 2: (a) and (b): Time and error comparison of iFEE (blue) vs. batch-mode (red) Fréchet mean computation for data from a symmetric distribution on $\mathbb{P}(6)$. (c) and (d): Similar experiment for data from an asymmetric distribution on $\mathbb{P}(4)$.

Asymmetric Distributions: For asymmetric distributions, we use a mixture of two log-Normal distributions on $\mathbb{P}(4)$ and repeat the same experiment as above. The first distribution in the mixture is centered at the identity matrix with the variance 0.1, and the second component is centered at a randomly-chosen matrix with variance 0.2. A set of 500 samples are drawn from this distribution for the experiment. To measure the error, we compute the gradient vector of the objective function in Equation 4 ($k = 500$) and its norm. Figure 2c depicts the timing plot for convergence of the two algorithms for samples drawn from this asymmetric distribution. The plot shown is an average over 500 runs. As evident, iFEE shows superior computational efficiency. Figure 2d depicts the error of iFEE vs. its counterpart batch mode algorithm for this asymmetric distribution. Note that the accuracy of both the algorithms are as expected similar.

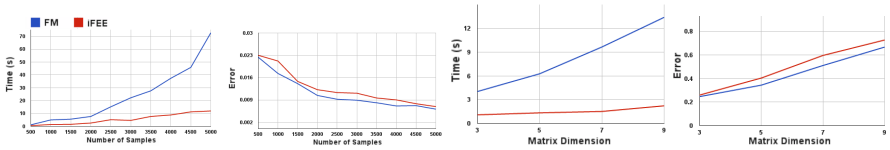
4.2 Application to K-means Clustering

In this section, we evaluate the performance of our proposed incremental algorithm within the K-means clustering framework. K-means clustering is of fundamental importance for many applications in computer vision and machine learning. Due to the lack of a closed-form formula for computing the Fréchet mean, mean computation is the most time consuming step in applying K-means to SPD matrices, since at the end of each iteration the mean for each estimated cluster needs to be recomputed. The experimental results in this section demonstrate that, for SPD matrices, our iFEE algorithm can significantly speed up the clustering process – when compared with the batch-mode – without any observable degradation in its accuracy.

For comparisons, we use the two different ways to compute the cluster centers: (i) the iFEE algorithm and (ii) the batch-mode algorithm for computing the Fréchet mean (FM). iFEE is applied to the K-means clustering for SPD matrices as follows. At the end of each iteration of the K-means algorithm, we only consider matrices whose cluster assignments have changed. For each of these "moving" samples, the source cluster center is updated by removing the sample, and the destination cluster center is updated by adding the new sample matrix. Both these updates can be effi-

ciently performed using our incremental formula given in Equation 14, with appropriate weights. A set of experiments are presented here using different scenarios to illustrate the effectiveness of our method. In each experiment, a set of random samples from mixtures of log-Normal distributions on $\mathbb{P}(n)$ are generated and used as inputs to the K-means algorithm. In the first experiment, we increase the number of samples and compare the accuracy and running time of incremental and batch-mode estimates for each case. In the second experiment, we evaluate the performance of each algorithm with respect to the matrix dimension. To measure the clustering error, the geodesic distance between each estimated cluster center and its true value is computed and these are summed over all cluster centers and reported.

Figures 3a and 3b, respectively compare the running time and the clustering accuracy of each method with increasing number of samples. It is evident that the iFEE outperforms the batch-mode method, while the accuracy for both methods are very similar. Moreover, as the number of samples increases, iFEE improves in accuracy. Also Figure 3c illustrates a significant difference in running time between these two methods, while Figure 3d shows that the accuracy for both methods are roughly the same. These experiments verify that the proposed iFEE is far more computationally efficient the batch-mode algorithm for K-means clustering applied to SPD matrices.



(a) (b) (c) (d)

Fig. 3: Comparison of running times and accuracy for K-means clustering based on iFEE and batch-mode estimators for, (a) and (b): varying number of samples from 3 clusters on $\mathbb{P}(4)$; (c) and (d): 1000 samples from 3 clusters with varying sizes.

4.3 Application to Diffusion Tensor Image Segmentation

In this section, we present results of applying our iFEE algorithm to real data segmentation specifically, the diffusion tensor image (DTI) segmentation problem. Diffusion tensors are symmetric positive definite (coefficient) matrices in a quadratic approximation to the diffusivity function characterizing the water molecule diffusion in sample tissue being imaged using a medical imaging technique called diffusion magnetic resonance imaging [54]. Standard MRI acquisition is modified via the application of diffusion sensitizing magnetic field gradients at each voxel of an image grid to acquire the signal along the applied magnetic field gradient directions.

One acquires these signals over a sphere of directions and at each voxel of an image grid, the diffusivity function is then approximated by a zero-mean local Gaussian, whose covariance matrix is proportional to the diffusion tensor. For more details on DTI, we refer the reader to [54].

In [14], the classical level-set based (piece-wise constant model) segmentation algorithm [42] was generalized to cope with a field of diffusion tensors. The constant in the piece-wise constant model employed here is the mean of the tensor-valued voxels in a region. In this section, we use this algorithm to segment DTIs, and use different (tensor field) mean estimation techniques within this algorithm for comparison purposes. We applied six different methods to compute the mean diffusion tensor (SPD matrices), and compared their accuracies and computation speed in the task of DTI segmentation. The first two methods used here are the proposed $iFEE$ and the batch-mode Fréchet mean (FM) obtained from Equation 4. The next two methods denoted henceforth by *KLS* and *RKLS* respectively are, the mean computed using the symmetrized KL-divergence [14] and its recursive counterpart, reported in [30]. The last two techniques are the Log-Euclidean mean (LEM) [55] and its recursive version ($RLEM$) introduced in [56].

The diffusion tensors at each grid point of the image field are estimated (using the method described in [14]) from a diffusion MR scan of a rat spinal cord. The data was acquired using a 17.6-T Brucker scanner, along 21 directions with a b-value of $1000 s/mm^2$. Each voxel size in the scan was $35\mu m \times 35\mu m \times 300\mu m$; and the image resolution was 128×128. Our goal here is to segment the gray matter from the DTI of the rat spinal cord. We used the same initialization for all the methods. We applied all of the six methods (incremental and batch-mode versions for each of the three "distance" measures) to perform this experiment. In order to compare the time efficiency, we report the run times for the entire gray matter segmentation process, including the total time required to compute the means. Table 1 shows the result of this comparison, from which we can see that FM takes nearly two thirds of the total reported segmentation time to compute the Fréchet mean, whereas, using the iFEE makes the computation much faster, and also significantly reduces the total segmentation time.

	$iFEE$	FM	$RKLS$	KLS	$RLEM$	LEM
MT	1.85	92.05	2.89	4.20	1.45	26.59
TT	52.42	147.79	54.23	59.41	66.34	117.58

Table 1: Time (in seconds) for segmentation of the gray matter in a rat spinal cord. MT and TT denote the mean computation time, and total segmentation time, respectively.

The segmentation results are shown in Figure 4 for each method. For the sake of space, we present only the segmentation results from $iFEE$, FM, *RKLS* and *RLEM* algorithms, as the results from *RKLS* and *RLEM* are visually similar to their non-incremental counterparts. The segmented region is the gray matter in the rat spinal cord. The region surrounding the entire spinal cord shown in blue is water in which the excised spinal cord was suspended for ex-vivo image acquisition. Each

$(3,3)$ diffusion tensor in the DTI data are depicted using an ellipsoid, whose axis directions and lengths correspond to the eigen-vectors and eigen-values of the tensor respectively. The color of each ellipsoid ranges from blue to red, demonstrating the lowest to highest degree of anisotropy respectively. Moreover, the segmentation result is depicted as a curve in red overlaid on the ellipsoidal visualization of the diffusion tensor field. From the figure, we can see that the segmentation results are visually similar to each other, while the iFEE takes far less computation time, which would be very useful in practice.

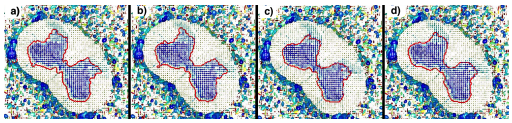


Fig. 4: Segmentation results of gray matter in a rat spinal cord. (a)-(d) Results from $iFEE$, FM, *RKLS* and *RLEM* respectively.

5 Conclusions

In this chapter, we have presented a novel incremental Fréchet expectation estimator dubbed iFEE that incrementally computes the Fréchet expectation of a distribution defined on a Riemannian manifold, and presented a proof for the algorithm's convergence for simply connected and complete Riemannian manifolds with non-positive sectional curvature. In iFEE, the intrinsic mean update is done by moving the current mean towards the new sample on the geodesic joining them; therefore, provided that the geodesics are accessible, iFEE does not require optimization and is computationally very efficient. The asymptotic accuracy of iFEE is guaranteed by the convergence analysis, and it provides an example of *geometric generalizations of the law of large numbers* in that the well-known sample average in the Euclidean law of large numbers is now replaced by the geometric operation of moves on geodesics. We have presented several experiments demonstrating the efficiency and accuracy of the iFEE algorithm.

Acknowledgement This research was supported in part by the NIH grant NS066340 to BCV.

References

1. Hartley, R., Trumpf, J., Dai, Y., Li, H.: Rotation averaging. Int. Journal of Computer Vision **103** (2013) 267–305
2. Kendall, D.: Shape manifolds, procrustean metrics, and complex projective spaces. Bulletin of the London Mathematical Society **16** (1984) 18–121
3. Subbarao, R., Meer, P.: Nonlinear mean shift over Riemannian manifolds. Int. Journal of Computer Vision **84** (2009) 1–20

4. Doretto, G., Chiuso, A., Wu, Y.N., Soatto, S.: Dynamic textures. Int. Journal of Computer Vision **51** (2003) 91–109
5. Turaga, P., Veeraraghavan, A., Srivastava, A., Chellappa, R.: Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence **33** (2011) 2273–2286
6. Cetingul, H., Vidal, R.: Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition. (2009) 1896–1902
7. Horn, B.: Robot Vision. The MIT Press (1886)
8. Tyagi, A., Davis, J.: A recursive filter for linear systems on Riemannian manifolds. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition. (2008)
9. Cheng, G., Vemuri, B.C.: A novel dynamic system in the space of SPD matrices with applications to appearance tracking. SIAM Journal on Imaging Sciences **6** (2013) 592–615
10. Tuzel, O., Porikli, F., Meer, P.: Human detection via classification on Riemannian manifolds. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition. (2007)
11. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. Int. Journal of Computer Vision **66** (2006) 41–66
12. Fletcher, P., Joshi, S.: Riemannian geometry for the statistical analysis of diffusion tensor data. Signal Processing **87** (2007) 250–262
13. Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. Journal of Mathematical Imaging and Vision **25** (2006) 423–444 10.1007/s10851-006-6897-z.
14. Wang, Z., Vemuri, B.: DTI segmentation using an information theoretic tensor dissimilarity measure. IEEE Trans. on Medical Imaging **24** (2005) 1267–1277
15. Amari, S.: Information Geometry. American Mathematical Society (2001)
16. Fréchet, M.: Les éléments aléatoires de nature quelconque dans un espace distancié. Annales de L'Institut Henri Poincaré **10** (1948) 215–310
17. Berger, M.: A Panoramic View of Riemannian Geometry. Springer (2007)
18. Sverdrup-Thygeson, H.: Strong law of large numbers for measures of central tendency and dispersion of random variables in compact metric spaces. ANN STAT **9** (1981) 141–145
19. Ziezold, H.: On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. Trans. of the Seventh Prague Conf. on Info. Theory, Statistical decision Functions, Random Processes and of the 1974 European Metting of Statisticians (1977)
20. Bhattacharya, R., Patrangenaru, V.: Large sample theory of intrinsic and extrinsic sample means on manifolds - II. The Annals of Statistics **33** (2005) 1225–1259
21. Bhattacharya, R., Patrangenaru, V.: Large sample theory of intrinsic and extrinsic sample means on manifolds - I. The Annals of Statistics **31** (2003) 1–29
22. Kendall, W., Le, H.: Limit theorems for empirical Frechet means of independent and non-identically distributed manifold-valued random variables. Brazilian Journal of Probability and Statistics **25** (2011) 323–352
23. Ginestet, C.: Strong and weak laws of large numbers for Frechet sample means in bounded metric spaces. arXiv:1204.3183 (2012)
24. Levy, A., Lindenbaum, M.: Sequential karhunenloeve basis extraction and its application to images. IEEE Trans. Image Processing **9** (2000) 1371–1374
25. Ross, D., Lim, J., Lin, R.S., Yang, M.H.: A Riemannian framework for tensor computing. Int. Journal of Computer Vision **77** (2008) 125–141
26. Cheeger, J., Ebin, D.: Comparison Theorems in Riemannian Geometry. American Mathematical Society (2008)
27. Khoshnevisan, D.: Probability. Volume 80 of Graduate Studies in Math. American Mathematical Society (2007)
28. Helgason, S.: Differantial Geometry, Lie Groups, and Symmetric Spaces. American Mathematical Society (2001)
29. Le, H.: Locating Féchet means with application to shape spaces. Advances in Applied Probability **33** (2001) 324–338
30. Cheng, G., Salehian, H., Vemuri, B.C.: Efficient recursive algorithms for computing the mean diffusion tensor and applications to DTI segmentation. In: ECCV. Volume 7. (2012) 390–401

31. Ho, J., Cheng, G., Salehian, H., Vemuri, B.: Recursive karcher expectation estimators and geometric law of large numbers. In: Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics. (2013) 325–332
32. Sturm, K.T.: Probability measures on metric spaces of nonpositive curvature. Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces **338** (2003)
33. Moakher, M.: A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM J. Matrix Analysis Applications **26** (2005) 735–747
34. Bhatia, R., Holbrook, J.: Riemannian geometry and matrix geometric means. Linear Algebra Appl. **413** (2006) 594–618
35. Fletcher, T., Joshi, S.: Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In: Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis. (2004) 87–98
36. T. Ando, CK Li, R.M.: Geometric means. Linear Algebra Appl. **385** (2004) 305–334
37. Bini, D.A., Meini, B., Poloni, F.: An effective matrix geometric mean satisfying the ando-li-mathias properties. Math. Comput. **79** (2010) 437–452
38. Lim, Y., Pálfia, M.: Weighted inductive means. Linear Algebra Appl. **453** (2014) 59–83
39. Pennec, X.: Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. Journal of Mathematical Imaging and Vision **25** (2006) 127–154
40. Afsari, B., Tron, R., Vidal, R.: On the convergence of gradient descent for finding the riemannian center of mass. SIAM Journal on Control and Optimization **51** (2013) 2230–2260
41. Wang, Z., Vemuri, B.: Tensor field segmentation using region based active contour model. In: European Conf. on Computer Vision (ECCV). (2004) 304–315
42. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. on Image Proc. **10** (2001) 266–277
43. Tsai, A., Yezzi, A.J., Willsky, A.: Curve Evolution Implementation of the Mumford-Shah Functional for Image Segmentation, Denoising, Interpolation, and Magnification. IEEE Trans. on Image Proc. **10** (2001) 1169–1186
44. Feddern, C., Weickert, J., Burgeth, B.: Level-set Methods for Tensor-valued Images. In: Proc. 2nd IEEE Workshop on Variational, Geometric and Level Set Methods in Comp. Vis. (2003) 65–72
45. Malladi, R., Sethian, J., Vemuri, B.C.: Shape Modeling with Front Propagation: A Level Set Approach. IEEE Trans. on PAMI **17** (1995) 158–175
46. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic Active Contours. Intl. Journ. of Compu. Vision **22** (1997) 61–79
47. Moakher, M., Batchelor, P.G. In: Symmetric Positive-Definite Matrices: From Geometry to Applications and Visualization. Visualization and Processing of Tensor Fields, Springer (2006)
48. Barmpoutis, A., Vemuri, B.C., Shepherd, T.M., Forder, J.R.: Tensor splines for interpolation and approximation of DT-MRI with applications to segmentation of isolated rat hippocampi. IEEE Trans. Med. Imag. **26** (2007)
49. Ziyan, U., Tuch, D., Westin, C.: Segmentation of thalamic nuclei from DTI using spectral clustering. MICCAI (2006) 807–814
50. Weldeselassie, Y., Hamarneh, G.: DT-MRI segmentation using graph cuts. In: SPIE Medical Imaging. Volume 6512. (2007)
51. Ho, J., Xie, Y., Vemuri, B.C.: On A nonlinear generalization of sparse coding and dictionary learning. In: ICML. (2013) 1480–1488
52. Xie, Y., Vemuri, B.C., Ho, J.: Statistical analysis of tensor fields. In: MICCAI. (2010) 682–689
53. Schwartzman, A.: Random ellipsoids and false discovery rates: Statistics for diffusion tensor imaging data. PhD thesis, Stanford University (2006)
54. Basser, P.J., Mattiello, J., Lebihan, D.: Estimation of the effective self-diffusion tensor from NMR spin echo. Journal of Magnetic Resonance **103** (1994) 247–254
55. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magn. Reson. Med. **56** (2006) 411–421
56. Wu, Y., Wang, J., Lu, H.: Real-time visual tracking via incremental covariance model update on log-euclidean Riemannian manifold. In: CCPR. (2009)
57. Bridson, M., Haefliger, A.: Metric Spaces of Non-Positive Curvature. Springer (1999)

APPENDIX

In this appendix, we prove Proposition-1 presented in Section two. The proof is entirely elementary if we assume the general property of $\mathbf{CAT}(0)$-metric spaces [57] and Toponogov's comparison theorem (specifically, the easier half of the theorem on manifolds of non-positive sectional curvature) [17].

Complete Riemannian manifolds of non-positive sectional curvature form an important subclass of $\mathbf{CAT}(0)$-metric spaces [57]. For our purpose, the detailed definition of $\mathbf{CAT}(0)$-metric spaces is not necessary; instead, we will recall only the features that are used in the proof. A geodesic triangle Γ on a complete Riemannian manifold $\mathscr{M}$ is the union of three geodesic segments joining three points $\mathbf{p},\mathbf{q},\mathbf{r} \in \mathscr{M}$: $\gamma_1(0) = \gamma_3(1) = \mathbf{p}, \gamma_1(1) = \gamma_2(0) = \mathbf{q}, \gamma_2(1) = \gamma_3(0) = \mathbf{r}$. Its comparison triangle Δ is a triangle in $\mathbb{R}^2$ with vertices p,q,r such that the lengths of the sides $\overline{pq}, \overline{qr}, \overline{rp}$ equal to the lengths of the geodesic segments $\gamma_1, \gamma_2, \gamma_3$, respectively. Such comparison triangle always exists for any geodesic triangle in $\mathscr{M}$, and it is unique up to a rigid transform in $\mathbb{R}^2$. The correspondence between the three sides and segments extends naturally to points on the triangles as well: a point $\mathbf{x} \in \Gamma$ corresponds to a point $x \in \Delta$ if their associated sides correspond and their distances to the corresponding two endpoints are the same. For example, if $\mathbf{x} \in \gamma_1$ and $x \in \overline{pq}$, $\mathbf{x}$ corresponds to x if $\mathbf{d}_{\mathscr{M}}(\mathbf{x},\mathbf{p}) = \mathbf{d}_{\mathbb{R}^2}(x,p)$, and hence $\mathbf{d}_{\mathscr{M}}(\mathbf{x},\mathbf{q}) = \mathbf{d}_{\mathbb{R}^2}(x,q)$ as well. An important property enjoyed by any $\mathbf{CAT}(0)$-metric space is that for any pair of points $\mathbf{x},\mathbf{y}$ on Γ and their corresponding points x,y on Δ, we have (see Figure 5) $\mathbf{d}_{\mathscr{M}}(\mathbf{x},\mathbf{y}) \leq \mathbf{d}_{\mathbb{R}^2}(x,y)$. The importance of this inequality is the upper bound given by the Euclidean distance in $\mathbb{R}^2$, and it allows us to bound the integral of the squared distance function on $\mathscr{M}$ by an integral involving squared Euclidean distance that is considerably easier to manage. Finally, for the pair of triangles Γ,Δ, Toponogov's comparison theorem asserts the angle $\angle(rpq)$ on Δ no smaller than $\angle(\mathbf{rpq})$ on Γ.

Armed with these results, the proof of Proposition-1 is straightforward and it involves comparing two triangles in the tangent space $\mathbf{T_m}$. See Figure 1. We restate proposition-1 for convenience below and now present its proof.

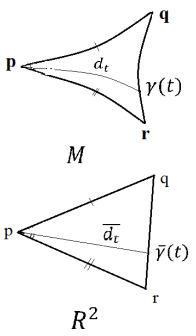


Fig. 5: A geodesic triangle in $\mathscr{M}$ and its comparison triangle in $\mathbb{R}^2$. Corresponding sides on the triangles have the same length. By Toponogov's comparison theorem, the angle $\angle(qpr)$ is not less than the angle $\angle(\mathbf{qpr})$ due the non-positive sectional curvature of $\mathscr{M}$. Furthermore, if $\gamma(t), \overline{\gamma}(t)$ denote the geodesic and straight line joining $\mathbf{p},\mathbf{q}$ and p,q, respectively, then the geodesic distance d_t between $\mathbf{p}$ and $\gamma(t)$ is not greater than the Euclidean distance $\overline{d_t}$ between p and $\overline{\gamma}(t)$, i.e., $d_t \leq \overline{d_t}$.

Proposition 1. *Let* $\mathbf{x},\mathbf{y},\mathbf{z}$ *be two points on a complete Riemannian manifold* $\mathscr{M}$ *with non-positive sectional curvature and* $\gamma(t)$ *the unique geodesic path joining* $\mathbf{x},\mathbf{y}$ *such that* $\gamma(0)=\mathbf{x},\gamma(1)=\mathbf{y}$. *Furthermore, let* $x=\mathbf{Log}_{\mathbf{z}}(\mathbf{x}),y=\mathbf{Log}_{\mathbf{z}}(\mathbf{y})$, *and* $\overline{\gamma}(t)$ *denote the straight line joining* x,y *such that* $\overline{\gamma(0)}=x,\overline{\gamma(1)}=y$. *Then,* $\mathbf{d}_{\mathscr{M}}(\gamma(t),\mathbf{z})\leq\|\overline{\gamma(t)}\|$.

Proof. Given $\mathbf{m}_k$, $\mathbf{x}_{k+1}$ in $\mathscr{M}$, and $\mathbf{m}_{k+1}$ determined according to the iFEE algorithm, we will denote m_k,x_{k+1} and m_{k+1}, their corresponding points in $\mathbf{T}_{\mathbf{m}}$ under the Riemannian logarithm map $\mathbf{Log}_{\mathbf{m}}$. Without loss of generality, we will prove the proposition using $\mathbf{z}=\mathbf{m}_k,\mathbf{x}=bx_{k+1},\mathbf{y}=\mathbf{m}_k$. Let $a=\mathbf{d}_{\mathscr{M}}(\mathbf{x}_{k+1},\mathbf{m})$ and $b=\mathbf{d}_{\mathscr{M}}(\mathbf{m}_k,\mathbf{m})$. On $\mathbf{T}_{\mathbf{m}}$, we have the first triangle Σ formed by the three vertices: x_{k+1},m_k and $\mathbf{o}$ the origin with the side lengths $|\overline{x_{k+1}\mathbf{o}}|=a,|\overline{m_k\mathbf{o}}|=b$. The geodesic triangle Γ on $\mathscr{M}$ spanned by $\mathbf{m},\mathbf{x}_{k+1},\mathbf{m}_k$ has a comparison triangle Δ in $\mathbf{T}_{\mathbf{m}}$ spanned by $\mathbf{o},p,q$ with $|\overline{p\mathbf{o}}|=a,|\overline{q\mathbf{o}}|=b$, and by Toponogov's comparison theorem,

$$\theta_\sigma\equiv\angle(x_{k+1}\mathbf{o}m_k)\leq\angle(p\mathbf{o}q)\equiv\theta_\delta,$$

since, by definition, $\angle(x_{k+1}\mathbf{o}m_k)=\angle(\mathbf{x}_{k+1}\mathbf{m}\mathbf{m}_k)$.

For completing the proof, we need to show that the distance between any point on the side $\overline{pq}$ of Δ and the origin is not greater than the distance between its corresponding point on the side $\overline{x_{k+1}m_k}$ of Σ and the origin. Specifically, a point $u\in\overline{pq}$ can be written as $u=tp+(1-t)q$, for some $0\leq t\leq 1$ and its corresponding point v on $\overline{x_{k+1}m_k}$ is the point $v=tx_{k+1}+(1-t)m_k$. Since the triangle Δ is unique up to a rigid transform, we can, without loss of generality, assume that the two triangles Δ,Σ are contained in a two-dimensional subspace of $\mathbf{T}_{\mathbf{m}}$ such that (using the obvious coordinates) they are spanned by the following two sets of three points:

$$\begin{aligned}\Delta:&\quad(0,0),(a,0),(b\cos\theta_\delta,b\sin\theta_\delta),\\ \Sigma:&\quad(0,0),(a,0),(b\cos\theta_\sigma,b\sin\theta_\sigma),\end{aligned}$$

with $\theta_\sigma\leq\theta_\delta$. Consequently, $u=(ta+(1-t)b\cos\theta_\delta,(1-t)b\sin\theta_\delta)$ and $v=(ta+(1-t)b\cos\theta_\sigma,(1-t)b\sin\theta_\sigma)$, and their lengths are, respectively,

$$\begin{aligned}\|u\|&=\sqrt{t^2a^2+2t(1-t)ab\cos\theta_\delta+(1-t)^2},\\ \|v\|&=\sqrt{t^2a^2+2t(1-t)ab\cos\theta_\sigma+(1-t)^2}.\end{aligned}$$

Since $\theta_\sigma\leq\theta_\delta$, it then follows that $\|u\|\leq\|v\|$.