Recursive Fréchet Mean Computation on the Grassmannian and its Applications to Computer Vision

Rudrasis Chakraborty and Baba C. Vemuri
Department of CISE, University of Florida
Gainesville, FL 32611, USA
{rudrasis, vemuri}@cise.ufl.edu

Abstract

In the past decade, Grassmann manifolds (Grassmannian) have been commonly used in mathematical formulations of many Computer Vision tasks. Averaging points on a Grassmann manifold is a very common operation in many applications including but not limited to, tracking, action recognition, video-face recognition, face recognition, etc. Computing the intrinsic/Fréchet mean (FM) of a set of points on the Grassmann can be cast as finding the global optimum (if it exists) of the sum of squared geodesic distances cost function. A common approach to solve this problem involves the use of the gradient descent method. An alternative way to compute the FM is to develop a recursive/inductive definition that does not involve optimizing the aforementioned cost function. In this paper, we propose one such computationally efficient algorithm called the Grassmann inductive Fréchet mean estimator (GiFME). In developing the recursive solution to find the FM of the given set of points, GiFME exploits the fact that there is a closed form solution to find the FM of two points on the Grassmann. In the limit as the number of samples tends to infinity, we prove that GiFME converges to the FM (this is called the weak consistency result on the Grassmann manifold). Further, for the finite sample case, in the limit as the number of sample paths (trials) goes to infinity, we show that GiFME converges to the finite sample FM. Moreover, we present a bound on the geodesic distance between the estimate from GiFME and the true FM. We present several experiments on synthetic and real data sets to demonstrate the performance of GiFME in comparison to the gradient descent based (batch mode) technique. Our goal in these applications is to demonstrate the computational advantage and achieve comparable accuracy to the state-of-the-art.

1. Introduction

The Grassmann manifold has been commonly used in mathematical formulations of many computer vision tasks such as video based face recognition [34, 4], activity recognition [5, 35], video restoration [15], shape analysis [13, 25] etc. Given a set of data points on a Grassmann manifold, finding their average is a crucial task in many classification and clustering based applications including the aforementioned applications. Since, a general Riemannian manifold lacks vector space structure, a Euclidean mean is not a valid representative of the average of points on the manifold. Instead one can use a Riemannian center of mass, also known as the Fréchet mean (FM) [12, 18] as a way to denote the average of data points. This mean however is not unique in general and can be shown to be unique only in a geodesic ball of a certain injectivity/convexity radius. We refer the reader to an excellent paper [2] for a detailed exposition on this topic. Further, in general there is no closed form solution for the FM of an arbitrary number of points on a Riemannian manifold. Of course, one can argue about using the Euclidean mean which though computationally efficient, is not intrinsic and does not have many of the desired properties of the FM on the manifold. On a Riemannian manifold, the FM is defined as the minimizer of the sum of squared geodesic distances cost function. A popular approach to compute the FM is to use a gradient descent technique on this cost function [10, 24, 3]. An alternative approach to finding the FM proposed in [14] however does not involve the minimization of a cost function but is posed as a special case of iteratively finding the zeros of a vector field on the Riemannian manifold.

Where applicable, a recursive algorithm can take advantage of the closed form solution to compute the FM of two points as the base of the recursion and recurse through the number of points in the given set. This will yield a much faster way to compute the FM if and when the convergence of the algorithm can be proved. In [31], an inductive definition of FM was proposed and almost sure convergence

was shown for all non-positively curved (NPC) spaces. In [33], authors also presented a recursive FM computation algorithm but without any convergence analysis. Recently, recursive algorithms (along with proofs of convergence) to compute the FM on the Riemannian manifold of $(n \times n)$ symmetric positive definite matrices (P_n) were reported in [16, 27, 21]. In this paper, we present a novel recursive (also called inductive) algorithm with convergence analysis, for computing the FM on a Grassmann manifold whose geometry is quite distinct from that of P_n and hence the convergence analysis on P_n [16, 27, 21] does not carry over to this setting.

The Grassmann manifold, denoted by $Grass(p, n)$, is the space of p dimensional subspaces of $\mathbf{R}^n$. In some of the recent works [34], researchers have used Procrustes metric defined in the ambient Euclidean space [7] instead of using an canonical metric on the Grassmann. As $Grass(p, n)$ is geodesically complete, hence, by the Hopf-Rinow theorem [17], there exists at least one length minimizing geodesic between any two points on $Grass(p, n)$. In our formulation, we use this closed form expression for geodesics derived using the unique (up to a multiplicative constant) unitary invariant Riemannian metric on the Grassmann [9, 1]. Note that in Euclidean space, the recursive form of computing the arithmetic mean (which yields the same solution as the minimization of sum of squared distances) involves only two points in each recursion step and can be geometrically interpreted as moving an appropriate distance away from the already computed mean (old mean) towards the new-mean on the straight line joining the old mean and the new data point. This geometric procedure can be readily extended to any Riemannian manifold using geodesics. To this end, we make use of the closed form expression – derived using the cannonical metric on the Grassman – for the geodesic between two points on the Grassman. More precisely, after computing the estimate of the FM of k points, denoted by $\mathcal{M}_k$, the $k + 1^{th}$ estimate lies on the geodesic between $\mathcal{M}_k$ and the $k + 1^{th}$ point $\mathcal{X}_{k+1}$. This readily yields an algorithm for computing the FM 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 minimization of sum of squared geodesic distances. However, because of the presence of curvature, the recursive form and the minimization formulation do not necessarily yield the same solution in general. In particular, it is not immediately clear that $\mathcal{M}_k$ will indeed converge asymptotically to the true FM.

In this paper, we present a weak consistency result which implies that as the number of samples goes to infinity, our inductive estimator does converge to the Fréchet expectation. For the finite samples case, with a fixed number of sample paths (trials), our estimator and FM in general can be different but close. Hence, we also give the bound on the distance between our estimator and the finite sample FM.

An excellent research monograph on the topic of statistics on special matrix manifolds including the Grassmann and Stiefel manifolds is the book by Chikuse [7]. The book however does not address computational efficiency issues. One of the first detailed and complete works on the gradient descent and the conjugate gradient descent algorithm on the Grassmann manifold was reported in [9]. Srivastava et al. [30] pose the commonly encountered Computer Vision problem of the subspace tracking problem on a complex Grassmann manifold. A nonlinear mean shift algorithm on the Grassmann manifold was presented in [32, 6] with applications to segmentation of multiple motions and filtering. Turaga et al. [34] have reported results on statistical analysis on the Grassmann manifold with various applications including action recognition, video face recognition etc. Most recently, Hauberg et al. [15] formulated the dimensionality reduction problem as an averaging problem on Grassmann. For a comprehensive survey of computer vision applications that use a variety of Riemannian manifolds including Grassmann we refer the reader to [23].

The rest of the paper is organized as follows. In section 2, we discuss the formulation of inductive FM on the Grassmann manifold. Further, we present a proof of the *Weak Consistency* of our estimator on the Grassmann. The distance between our inductive estimate of the FM and the true FM is also presented for the finite sample case. Experimental results on synthetic and real data sets are presented in section 3. Finally, in section 4, we draw conclusions.

2. Inductive Fréchet mean on the Grassmann manifold

In this section, we first briefly recall some preliminaries regarding the Grassmann manifold. Then, we present an inductive formulation for computing the Fréchet mean (FM) [12, 18] on Grassmann manifold. For a finite sample set, the FM is defined as a minimizer of the sum of squared geodesic distances between the unknown mean and the given samples. In general, the uniqueness can only be guaranteed within a certain injectivity/ convexity radius [12, 18, 2]. We also prove the *Weak Consistency* of GiFME on the Grassmannian and present a bound on the distance between the the estimate obtained using GiFME and the FM.

2.1. Mathematical Preliminaries

The set of all $p-$dimensional linear subspaces of $\mathbf{R}^n$ is defined as the Grassmann manifold, $Grass(p, n)$, where $p \in \mathbf{Z}^+$, $n \in \mathbf{Z}^+, n \geq p$. So, a point on $Grass(p, n)$ is a $p-$dimensional hyperplane in $\mathbf{R}^n$ containing the origin. The special case of $Grass(p, n)$ is when $p = 1$, which is called the *real projective space*. A point, $\mathcal{X} \in Grass(p, n)$

can be specified by a basis, i.e., a set of p, $n-$dimensional linearly independent vectors. Let $X_{n \times p}$ be the matrix whose columns are the basis vectors. Then, $\mathcal{X} = Col(X)$, where $Col(X)$ returns the column space of X. The set of all $n \times p$, $p \leq n$, full rank matrices is termed as (noncompact) Stiefel manifold, denoted by $St(p,n)$.

Clearly, given $\mathcal{X} \in Grass(p,n)$, the choice of $X \in St(p,n)$ for which $\mathcal{X} = Col(X)$ is not unique. In fact, $Grass(p,n)$ can be identified with the quotient space $St(p,n)/GL(p)$, where $GL(p)$ is the $p-$dimensional general linear group, i.e., $p \times p$ non-singular matrices. Thus, $GL(p)$ defines an equivalence relation on $St(p,n)$ as follows: $X \sim Y$ iff $\exists L \in GL(p)$ such that, $X = YL$, $X, Y \in St(p,n)$. Note that, $Col(X)$ is same as $Col(Y)$, so the equivalence classes of $St(p,n)$ are in one-to-one correspondence with the points on $Grass(p,n)$.

Now, we will consider the unitary group invariant canonical metric on $Grass(p,n)$. The terminology used here is borrowed from [1]. In order to define a metric on $Grass(p,n)$, we will first define a metric on $St(p,n)$. Let $\mathcal{X} \in Grass(p,n)$, $\xi, \eta \in T_{\mathcal{X}}Grass(p,n)$. Let, $X \in St(p,n)$ such that $\mathcal{X} = Col(X)$. As $St(p,n)$ is an open subset of $\mathbf{R}^{n \times p}$, $T_X St(p,n)$ is isomorphic to $\mathbf{R}^{n \times p}$. We define the *horizontal space* H_X (a subspace of $T_X St(p,n)$) as

$$H_X = \{X \perp W | W \in \mathbf{R}^{(n-p) \times p}\}$$

where $X \perp \in St(n-p,n)$ s.t. $X^T X \perp = 0$. For any two $U, V \in T_X St(p,n)$ define the following metric:

$$\langle U, V \rangle_X = trace\Big((X^T X)^{-1} U^T V\Big) \quad (1)$$

According to the theory of principal fiber bundles [19], for every $\xi \in T_{\mathcal{X}}Grass(p,n)$, there exists an unique horizontal vector $\xi_{\diamond X}$ such that $\xi_{\diamond X}$ projects to ξ via the span operation. Hence, using Eq. 1, we can define the metric on $Grass(p,n)$ as follows:

$$\langle \xi, \eta \rangle_{\mathcal{X}} = trace\Big((X^T X)^{-1} \xi_{\diamond X}^T \eta_{\diamond X}\Big) \quad (2)$$

where, $\xi_{\diamond X}$ denotes the *horizontal lift of ξ at X* [1]. This metric is invariant under the orthogonal group (set of $n \times n$ dimensional orthogonal matrices) operation, i.e.,

$$trace\Big((X^T X)^{-1} \xi_{\diamond X}^T \eta_{\diamond X}\Big)$$
$$= trace\Big(((QX)^T(QX))^{-1}(Q\xi_{\diamond X})^T(Q\eta_{\diamond X})\Big)$$

for all $Q \in O(n)$. Further, it is well known that the Grassmann manifold is complete [1] and hence, one can extend the geodesics on it indefinitely. Given $\mathcal{X}, \mathcal{Y} \in Grass(p,n)$ with their respective orthonormal basis (o.n.b.) X and Y, the unique geodesic, $\Gamma(\mathcal{X}, \mathcal{Y}, t)$ from $\mathcal{X}$ to $\mathcal{Y}$ is defined by [1]:

$$\Gamma(\mathcal{X}, \mathcal{Y}, t) = span\Big(XV \cos(\Theta t) + U \sin(\Theta t)\Big) \quad (3)$$

where, $X^T Y$ is non-singular and $\Big(I - X(X^T X)^{-1} X^T\Big) Y(X^T Y)^{-1} = U\Sigma V^T$ be the "thin" Singular value decomposition (SVD), (i.e., U is $n \times p$ and V is $p \times p$ orthonormal matrix, whereas Σ is $p \times p$ diagonal matrix), and $\Theta = atan\Sigma$. The geodesic distance between $\mathcal{X}$ and $\mathcal{Y}$ induced by the unitary invariant metric (as given in Eq. 2), is

$$d_{Gr}(\mathcal{X}, \mathcal{Y}) = \sqrt{\langle \xi, \xi \rangle} \quad (4)$$

where, $\xi = Exp_{\mathcal{X}}^{-1} \mathcal{Y}$, $Exp_{\mathcal{X}}^{-1} \mathcal{Y}$ is the inverse of the Riemannian Exponential map. The distance between $\mathcal{X}$ and $\mathcal{Y}$ can also be shown as the $2-$norm of $diag(\Theta)$, where, $diag(\Theta)$ returns the vector consisting of the diagonal elements of Θ (termed as principal angles between $\mathcal{X}$ and $\mathcal{Y}$). Another classical definition of distance on Grassmannian is given by $\sin(\max \Theta)$, where $\max \Theta$ is the largest principal angle. In this work, we have used the $\sin(\max \Theta)$ definition. We are now ready to use the above material to derive the incremental a.k.a inductive FM.

2.2. The Inductive Fréchet Expectation

Given the closed-form expression for the geodesic on a Grassmann manifold, now we are in a position to define the inductive Fréchet expectation estimator on the Grassmann manifold, henceforth abbreviated as $GiFEE$. Prior to that, let us recall the definition of the Fréchet Mean (FM) and the Fréchet Expectation. The FM, $\mathcal{X}^*$ of a set of N points $\{\mathcal{X}_1, \cdots, \mathcal{X}_N\}$, is defined [12, 18] as:

$$\mathcal{X}^* = \arg\min_{\mathcal{Y} \in Grass(p,n)} \sum_{i=1}^{N} d_{Gr}^2\Big(\mathcal{Y}, \mathcal{X}_i\Big) \quad (5)$$

For a probability density $f_{\mathcal{X}}$ on $Grass(p,n)$ with finite Riemannian $\mathbf{L}^2$ moment, the Fréchet Expectation, μ^* is defined [12, 18] as:

$$\mu^* = \arg\min_{\mu \in Grass(p,n)} \int_{Grass(p,n)} d_{Gr}^2\Big(\mu, \mathcal{X}\Big) f_{\mathcal{X}}(\mathcal{X}) d(\mathcal{X})$$
$$\quad (6)$$

where, $d(\mathcal{X})$ is the Riemannian measure (or volume form) used in defining the density on $Grass(p,n)$.

Algorithm for Inductive Fréchet Expectation Estimator Let $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$ be independent samples drawn from a probability distribution $P(\mathcal{X})$ on $Grass(p,n)$. Then, we define the Fréchet expectation estimator M_k by the following recursion:

$$M_1 = \mathcal{X}_1 \quad (7)$$
$$M_{k+1} = \Gamma\Big(M_k, \mathcal{X}_{k+1}, \omega_{k+1}\Big) \quad (8)$$

where, $\Gamma(.,.,.)$ is the geodesic as defined in Eq. 3, and

$\omega_{k+1} = \frac{1}{k+1}$. Eq. 7 simply means that the $k+1^{th}$ estimator lies on the geodesic between the kth estimator and the $k+1^{th}$ sample point. The rationale behind this formulation comes from the euclidean update formula as discussed in [16]. This inductive formulation will henceforth be denoted by $GiFEE$ for expectation estimator and $GiFME$ for the finite sample mean estimator respectively.

This simple inductive estimator can be shown to converge to the Fréchet expectation as stated in Theorem 1. The proof of this theorem is given in the next subsection.

Theorem 1. (Weak Consistency)*: For any distribution $P(\mathcal{X})$ on $Grass(p, n)$ with finite Riemannian $\mathbf{L}^2$-moment, whose support is bounded by a geodesic ball of radius $< \pi/4$, and a sequence of independent samples $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$ drawn from this distribution, the inductive Fréchet expectation estimator (GiFEE) converges to the Fréchet expectation in probability as the number of samples tends to ∞.*

2.3. Proof of Theorem 1

In order to prove Theorem 1, i.e., the *Weak Consistency of GiFEE on $Grass(p, n)$*, first we will prove three lemmas, which will be used to prove the Theorem 1. *Note that these proof steps are not a part of our inductive algorithm. These steps are only needed to prove the weak consistency on $Grass(p, n)$. Also, note that in the rest of paper with a slight abuse of notation, we have used the term "geodesic" to denote the "shortest geodesic".*

Lemma 1. *Let $\mathcal{X}$ and $\mathcal{Y}$ belong to $Grass(p, n), n/2 > p, p > 1$ and let X and Y be an orthonormal basis (o.n.b.) of $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let the geodesic between $\mathcal{X}$ and $\mathcal{Y}$ be denoted by $\Gamma(\mathcal{X}, \mathcal{Y}, t)$. Let $A \in M_{n \times (n-1)}(\mathbb{R})$ be of full column rank such that $A^T A = I_{n-1}$ and $A^T X$ is of rank p. Then, $\exists \widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}} \in Grass(p, n-1)$ with their respective o.n.b. $\widetilde{X}$ and $\widetilde{Y}$ such that $\Gamma(\mathcal{X}, \mathcal{Y}, t) = span(A\Gamma_\beta(t))$ where $\Gamma_\beta(t)$ is a basis of $\Gamma(\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}, t)$.*

Proof. : Since the distribution $P(\mathcal{X})$ has a support bounded by a geodesic ball of radius $< \pi/4$, the largest principal angle between $\mathcal{X}$ and $\mathcal{Y}$ is $< \pi/2$, i.e., $\mathcal{X}$ does not contain an orthogonal direction to $\mathcal{Y}$, which implies $X^T Y$ is invertible.

Claim: Given the hypothesis as above, $A^T Y$ is of full column rank.

Proof. : Consider the matrix $R = X^T A A^T Y$. Assume R is not of full rank, then $\exists \{c_i\}_{i=1}^p$, not all 0 s.t.,

$$\sum_i c_i x_i^t (A A^T) Y = 0 \tag{9}$$

where, x_i is the ith column of X. As Y and AA^T is not zero matrix, so Eq. 9 holds if either $\sum_i c_i x_i = 0$

or $\sum_i c_i x_i \in \mathcal{N}(AA^T)$. As X is of full column rank, $\sum_i c_i x_i = 0$ is true only if $c_i = 0$, $\forall i$, which is a contradiction to our assumption that R is not of full rank. Hence, $\sum_i c_i x_i \in \mathcal{N}(AA^T)$, $\implies AA^T \sum_i (c_i x_i) = 0$, $\implies A^T X$ is not of full column rank, which is a contradiction to our hypothesis. Hence, $R = X^T A A^T Y$ is invertible, $\implies A^T Y$ is of full column rank since, by hypothesis $A^T X$ is of full column rank. ∎

Now, construct $A \in M_{n \times (n-1)}(\mathbb{R})$ as follows. Make the first p columns of A same as those of X. Then, we take each column of Y individually, and use Gram-Schmidt orthogonalization on that column. A is then augmented with this orthogonalized column. Note that, if a column of Y is in the columnspace of X, then, it is not used in the augmentation of A. After these operations, we fill the remaning columns of A by taking o.n.b. of $\mathbb{R}^n$ s.t. they together make A column orthonormal. Clearly, as the first p columns of A are all the columns of X, $A^T X$ is of full column rank. Hence, by the above claim, $A^T Y$ is of full column rank. Let, $N \in M_{p \times p}(\mathbb{R})$ be invertible matrix such that

$$X = A\widetilde{X} \tag{10}$$

$$Y = A\widetilde{Y}N \tag{11}$$

where, $\widetilde{X}, \widetilde{Y} \in M_{(n-1) \times p}(\mathbb{R})$ are column orthonormal. Hence, $\exists \widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}} \in Grass(p, n-1)$ s.t. $\widetilde{X}, \widetilde{Y}$ are o.n.b. of $\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}$ respectively. It's easy to see from our construction of A that $\widetilde{X} = \begin{bmatrix} I_p \\ 0_{(n-1) \times p} \end{bmatrix}$. Moreover, as $Y \in Col(A)$, $\widetilde{Y}N$ gives the coefficient of the linear combination. Hence, we can use thin QR decomposition to get column orthonormal $\widetilde{Y}$ and the upper-triangular N with a non-zero diagonal.

Claim: $\widetilde{X}^T \widetilde{Y}$ is invertible.

Proof. : Using Equations 10 and 11, we get, $\widetilde{X}^T \widetilde{Y} = X^T A A^T Y N^{-1}$. By our earlier claim, the square matrix, $X^T A A^T Y$ is invertible. As product of invertible matrices is invertible, hence $\widetilde{X}^T \widetilde{Y}$ is invertible. ∎

Let $Y(X^T Y)^{-1} - X = U\Sigma V^T$ be the SVD decomposition. Using the previous equalities and Equations 10 and 11, we get

$$
\begin{aligned}
Y(X^T Y)^{-1} - X &= A\widetilde{Y}N(\widetilde{X}^T A^T A\widetilde{Y}N)^{-1} - A\widetilde{X} \\
&= A\big(\widetilde{Y}N(\widetilde{X}^T \widetilde{Y}N)^{-1} - \widetilde{X}\big) \quad (12) \\
&= A\big(\widetilde{Y}NN^{-1}(\widetilde{X}^T \widetilde{Y})^{-1} - \widetilde{X}\big) (13) \\
&= A\big(\widetilde{Y}(\widetilde{X}^T \widetilde{Y})^{-1} - \widetilde{X}\big) \quad (14)
\end{aligned}
$$

Let $\widetilde{Y}(\widetilde{X}^T \widetilde{Y})^{-1} - \widetilde{X} = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^T$. Then, using Equation 14, we obtain the relationship between (U, Σ, V) and $(\widetilde{U}, \widetilde{\Sigma}, \widetilde{V})$ as follows: $U = A\widetilde{U}$, $V = \widetilde{V}$ and $\Sigma = \widetilde{\Sigma}$. Let

$\Theta = \text{atan} \Sigma$ and $\widetilde{\Theta} = \text{atan} \widetilde{\Sigma} = \Theta$. The geodesic between $\mathcal{X}$ and $\mathcal{Y}$, $\Gamma(\mathcal{X}, \mathcal{Y}, t)$ can be expressed as [1]

$$\Gamma(\mathcal{X}, \mathcal{Y}, t) = span(\bar{X} V \cos \Theta t + U \sin \Theta t) \quad (15)$$

Substituting the relationship between (U, Σ, V) and $(\widetilde{U}, \widetilde{\Sigma}, \widetilde{V})$ into Equation 15, we get

$$\begin{aligned}
\Gamma(\mathcal{X}, \mathcal{Y}, t) &= span(\bar{X} V \cos \Theta t + U \sin \Theta t) \\
&= span(A[\widetilde{X} \widetilde{V} \cos \widetilde{\Theta} t + \widetilde{U} \sin \widetilde{\Theta} t]) \\
&= span(A \Gamma_\beta(t)) \quad (16)
\end{aligned}$$

where, $\Gamma_\beta(t)$ is the basis for $\Gamma(\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}, t)$. Thus, the geodesic between $\mathcal{X}$ and $\mathcal{Y}$ in $Grass(p, n)$ can be expressed as a function of $\Gamma_\beta(t)$. Note that $A \Gamma_\beta(t)$ is a basis for $\Gamma(\mathcal{X}, \mathcal{Y}, t)$, and $A^T A \Gamma_\beta(t) = \Gamma_\beta(t)$ is of full column rank. ∎

For, $Grass(p, n)$ with $p = n/2$, if X and Y together spans $\mathbb{R}^n$, $\nexists\, A$ in Lemma 1. In this special case, instead of the reduction from $Grass(p, n)$ to $Grass(p, n-1)$, we map from $Grass(p, n)$ to $Grass(p+1, n)$. In this case, we look for a $\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}} \in Grass(p+1, n)$ s.t. $d_{Gr}(\mathcal{X}, \mathcal{Y}) = d_{Gr}(\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}})$. So, in this case, we will get $X = \widetilde{X} B$ and $Y = \widetilde{Y} C$, where B and C contain the coefficient of the linear combination of columns of $\widetilde{X}$ and $\widetilde{Y}$ respectively. Then, we can use analogous arguments as in Lemma 1 to get the geodesic expression between $\mathcal{X}$ and $\mathcal{Y}$ from the geodesic expression between $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}}$. The detailed proof will be given in a future publication.

Lemma 2. *Let $\mathcal{X}$ and $\mathcal{Y}$ belong to $Grass(1, n)$ from a distribution whose support is bounded by geodesic ball of radius $< \pi/4$, and let $X, Y \in S^{n-1}$ be the unit length basis of $\mathcal{X}$ and $\mathcal{Y}$ respectively s.t. $d_S(X, Y) < \pi/2$, where $d_S(.,.)$ is the geodesic distance on S^{n-1}. Let $\Gamma_S(.,.,t)$ denote the geodesic on S^{n-1}. Then, any point on the geodesic, at $(t = t_0)$, between X and Y, i.e., $\Gamma_S(X, Y, t_0)$ spans $\Gamma(\mathcal{X}, \mathcal{Y}, t_0)$.*

Proof. Since X and Y are unit length basis, they lie on the hypersphere, S^{n-1}. Let $Y(X^T Y)^{-1} - X = U \Sigma V^T$ be the SVD decomposition. Then, $U \Sigma V^T = Y(X^T Y)^{-1} - X = \frac{Y - X \cos \theta}{\cos \theta}$, where, $X^T Y = \cos \theta$ and as $X^T Y$ is invertible thus, $\cos \theta \neq 0$. Then,

$$U = \frac{Y - X \cos \theta}{\|Y - X \cos \theta\|} \quad (17)$$

$$\Sigma = \frac{\|Y - X \cos \theta\|}{\cos \theta}$$
$$= \tan \theta \quad (18)$$
$$V = 1 \quad (19)$$

So, $\text{atan} \Sigma = \theta$. Using Equations 17-19, the geodesic $\Gamma(\mathcal{X}, \mathcal{Y}, t)$ between $\mathcal{X}$ and $\mathcal{Y}$, can be expressed as,

$$\begin{aligned}
\Gamma(\mathcal{X}, \mathcal{Y}, t) &= span(XV \cos \theta t + U \sin \theta t) \\
&= span(X \cos \theta t + \frac{Y - X \cos \theta}{\|Y - X \cos \theta\|} \sin \theta t) \\
&= span(\frac{1}{\sin \theta}[Y \sin \theta t + X \sin \theta(1 - t)]).
\end{aligned}$$
$$(20)$$

Since $X, Y \in S^{n-1}$, hence, Equation 20 implies that $\Gamma(\mathcal{X}, \mathcal{Y}, t) = span(\Gamma_S(X, Y, t))$. Here, as $\Gamma_S(X, Y, t) \in S^{n-1}$, hence, this becomes an unit length basis for the subspace on the geodesic between $\mathcal{X}$ and $\mathcal{Y}$. ∎

Lemma 3. *Given the hypothesis as above, independent samples on $Grass(p, n)$ induces independent samples (from the induced distribution) on $S^{n'-1}$ for some $n' \leq n$.*

Proof. Consider a set of independent samples on $Grass(p, n)$, $\mathcal{X}_1, \cdots, \mathcal{X}_N$. Using the proof of Lemma 1 and the claim within, we can take a full column orthonormal matrix A, to get samples $\widetilde{\mathcal{X}}_1, \cdots, \widetilde{\mathcal{X}}_N$ on $Grass(p, n-1)$. Note that since any functional tranformation of indepedent samples yields independent samples (from the induced distribution), $\widetilde{\mathcal{X}}_1, \cdots, \widetilde{\mathcal{X}}_N$ are independent samples. For the case when we need to go from $Grass(p, n)$ to $Grass(p+1, n)$, by analogous argument, the samples are also independent. So, we will use the isometry between $Grass(p, n)$ and $Grass(n - p, n)$ [7], the reduction step from $Grass(p, n)$ from $Grass(p, n-1)$ and the step from $Grass(p, n)$ from $Grass(p+1, n)$ to get independent samples on $Grass(1, n')$, for some $n' \leq n$. Now, define a mapping $\mathfrak{F} : Grass(1, n') \rightarrow S^{n'-1}$ by $\mathfrak{F}(\mathcal{X}) = X * sgn(X_n)/\|X\|$, where X is a basis of $\mathcal{X} \in Grass(1, n')$. Then, $\mathfrak{F}$ induces independent samples on $S^{n'-1}$ by the similar argument as before. ∎

Given the above lemmas, we are now in a position to prove Theorem 1.

Proof of Theorem 1. Given, Lemma 3, we can state that independent samples on $Grass(p, n)$ induces independent samples on $S^{n'-1}$. Using Lemma 2, we can state that the geodesic on $Grass(1, n)$ (between two points) can be computed from geodesic on S^{n-1} (between two corresponding points). And as our inductive FM computation algorithm, $GiFEE$, (in eqns. 7) requires closed form expression of the geodesic at each step, we can compute $GiFEE$ of the samples on $Grass(1, n)$ from the $SiFEE$ on S^{n-1}. Moreover, using Lemma 1, we can also compute $GiFEE$ on $Grass(p, n)$ from the $GiFEE$ on $Grass(p, n-1)$ as geodesic on $Grass(p, n)$ can be computed from the geodesic on $Grass(p, n-1)$. For $p = n/2$ case, we will

use the step to go from $Grass(p, n)$ to $Grass(p + 1, n)$. Additionally, we use the isometry between $Grass(p, n)$ and $Grass(n - p, n)$, $n - p > 0$ [7], to derive $GiFEE$ on $Grass(n - p, n)$ from the $GiFEE$ on $Grass(p, n)$. Hence, we can state that the $GiFEE$ of the samples on $Grass(p, n)$ can be derived from the $SiFEE$ on $S^{n'-1}$, for some $n' \leq n$. Hence, given the proof of weak consistency on $S^{n'-1}$, we have a proof for the weak consistency of GiFEE on $Grass(p, n)$.

An example of the reduction steps from $Grass(5, 10)$ **to** $Grass(1, 4)$: $Grass(5, 10) \rightarrow Grass(6, 10) \rightarrow Grass(4, 10) \rightarrow Grass(4, 9) \rightarrow Grass(4, 8) \rightarrow Grass(5, 8) \rightarrow Grass(3, 8) \rightarrow Grass(3, 7) \rightarrow Grass(3, 6) \rightarrow Grass(4, 6) \rightarrow Grass(2, 6) \rightarrow Grass(2, 5) \rightarrow Grass(2, 4) \rightarrow Grass(3, 4) \rightarrow Grass(1, 4)$, where $\rightarrow$ denotes the reduction step. *We now state the weak consistency of our estimator on S^{n-1} in Theorem 2 and refer the reader to [26] for the proof.* ∎

Theorem 2. (Weak consistency on S^{n-1}) *Let X_i's be the independent samples drawn from a distribution whose support is bounded by a geodesic ball of radius $< \pi/4$. Then the inductive Fréchet expectation estimator ($SiFEE$) will converge to the Fréchet estimator in probability as the number of samples tend to ∞.*

In the next subsection, we will show convergence for finite sample set.

2.4. The Finite Sample Case

Instead of infinite sample set, if we have finite sample set on $Grass(p, n)$, it can also be proved that, if we draw an infinite number of times from the finite sample set, $GiFME$ converges to the FM of the finite samples. This result is analogous to Sturm's result for non-positively curved (NPC) spaces [31] and proved for weighted inductive means on the manifold of symmetric positive definite matrices in Lim et. al [21].

Let $\mathbb{X} = \{\mathcal{X}_1, \cdots, \mathcal{X}_n\}$ be the set of n points on $Grass(p, n)$. Also, assume on $\mathbb{X}$, FM exists and it is unique. Let, $\mathbb{N}_n = \{1, \cdots, n\}$. For each $\omega \in \prod_{k=1}^{\infty} \mathbb{N}_n$, define inductively a sequence σ_ω as follows:

$$\sigma_\omega(1) = \mathcal{X}_{\omega(1)}$$

$$\sigma_\omega(m) = \Gamma\left(\sigma_\omega(m - 1), \mathcal{X}_{\omega(m)}, \frac{1}{m}\right)$$

This sequence may be viewed as a random walk starting at $\mathcal{X}_{\omega(1)}$, and moving from $\sigma_\omega(m - 1)$ toward $\mathcal{X}_{\omega(m)}$ along the geodesic to reach $\sigma_\omega(m)$. Then, the following theorem holds.

Theorem 3. (Convergence of the inductive random walk) *Given the hypothesis as above, as $m \rightarrow \infty$, $\sigma_\omega(m)$ converges almost surely to $\mathcal{X}^*$, i.e., the FM of $\mathbb{X}$. More concisely,*

$$\lim_{m \rightarrow \infty} \sigma_\omega(m) = \mathcal{X}^* \quad (21)$$

Proof. Let us construct a set $\mathbb{Y}_m = \{\mathcal{X}_{\omega(1)}, \cdots, \mathcal{X}_{\omega(m)}\}$. From Theorem 1, we can state that in the limit as m goes to ∞, $\sigma_\omega(m)$ converges to the FM of $\mathbb{Y}_m$, which is same as the FM of the set $\mathbb{X}$. ∎

The more generalized form of Theorem 3, where associated with $\mathbb{X}$ there is a probability vector $\mathbb{W} = (w_1, w_2, \cdots, w_n)$, is know as "no-dice theorem", which is harder to prove. In our case, $w_i = \frac{1}{n}, \forall i$.

In the following subsection, we will present a bound on the distance between $GiFME$ and the finite sample FM.

2.5. An upper bound on $d_{Gr}\left(M_{k+1}, \mathcal{X}^*\right)$

Given a finite number of sample points on $Grass(p, n)$, let the kth estimate from $GiFME$ be M_k and the FM be denoted by $\mathcal{X}^*$. Using Lemma 1 and the isometry between $Grass(p, n)$ and $Grass(n-p, n)$, it is easy to see that finding an upper bound on $d_{Gr}\left(M_{k+1}, \mathcal{X}^*\right)$ is equivalent to finding an upper bound on the distance between $SiFME$ and FM on S^{n-1}.

Lemma 4. *Given $\mathcal{S}_1, \cdots, \mathcal{S}_n$ on northern hemisphere of S^{n-1} (i.e., n^{th} coordinate of $\mathcal{S}_i$ is always positive, $\forall i$) with FM $\mathcal{S}^* = (0, 0, \cdots, 0, 1)^t$ and k^{th} inductive estimator, T_k, the distance between T_k and $\mathcal{S}^*$, denoted by $d_S(T_k, \mathcal{S}^*)$ is bounded according to,*

$$d_S(T_k, \mathcal{S}^*) \leq \max_i \theta_1^i$$

where, $\theta_1^i, \theta_2^i, \cdots, \theta_{n-1}^i$ are the angular coordinates for $\mathcal{S}_i$, $\theta_j^i \in (0, \pi), \forall \{i, j\}$.

The proof of this Lemma is given in [26].

3. Experimental Results

In this section, we present experiments demonstrating the performance of GiFME in comparison to the batch mode counterpart (which uses the gradient descent on the sum of squared geodesic distances cost function) on synthetic and real data. Since the practical thrust of the paper is on computational efficiency in estimating the FM using GiFME, we provide the run times on a desktop used for performing all the experiments. All the experimental results reported here were executed on a desktop with a single 3.33 GHz Intel-i7 CPU with 24 GB RAM.
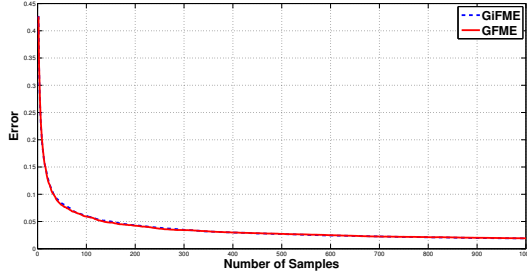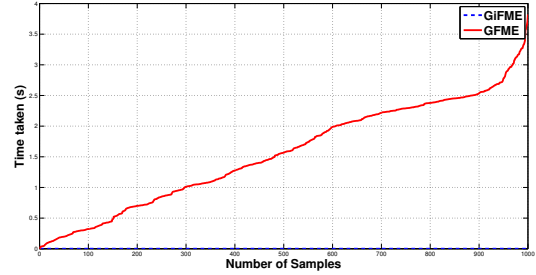
Figure 1. Average error comparison for $GFME$ and $GiFME$

Figure 2. The running time comparison for $GFME$ and $GiFME$ (in an incremental context).

Figure 3. The comparison of number of samples required for $GFME$ and $GiFME$ to attain a specified accuracy.

3.1. Comparative performance of GiFME on Synthetic data

We generate 1000 i.i.d. samples from a Log-Normal distribution on $Grass(p, n)$ with variance 0.25 and expectation to be $Col(\tilde{I})$, where

$$\tilde{I}_{ij} = \begin{cases} 1 & 1 \leq i = j \leq p \\ 0 & \text{o.w.} \end{cases}$$

We use the procedure in [29] to generate samples form a Log-Normal distribution. Then, we input these i.i.d. samples to both $GiFME$ and $GFME$ (non-recursive, Grassman Fréchet mean estimator). To compare the performance we compute the error, which is the intrinsic distance (on the Grassmann) between the computed mean and the FM. We also report the computation time for these both cases. We perform this experiment 10 times and the average error in accuracy and the average computation time are reported. The comparison plot of the average error is shown in Fig. 1, here $n = 100$, $p = 10$. In order to achieve faster convergence of $GFME$, we have used the FM of k samples to initialize the FM computation for $k + 1$ samples. From this plot, it can be seen that the average accuracy error of $GiFME$ is almost same as that of $GFME$. The comparison plot of computation time for $GiFME$ and $GFME$ is shown in Fig. 2. In Fig. 2 we present the time required by $GiFME$ for the incremental update step involved in computing the FM of k samples given the FM of $(k - 1)$ samples. From this figure, we can see that $GiFME$ outperforms $GFME$. As the number of samples increases, the computational efficiency of $GiFME$ over $GFME$ becomes very large. We can also see that the time requirement for $GiFME$ is almost constant with respect to the number of samples, which makes $GiFME$ computationally very efficient and attractive for large number of samples.

Another interesting question to ask is, how many samples are need in order to compute the FM within a given error tolerance? We answer this question through the plot in Figure 3 and present a comparison of the number of sample required for $GiFM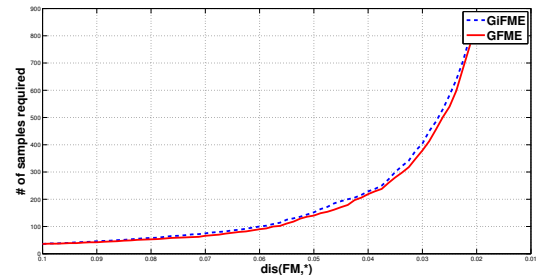E$ and $GFME$ to reach the specified error tolerance. From Fig. 3, we can see that the number of samples required to reach the specified error tolerance by $GiFME$ and $GFME$ are almost the same.

3.2. Application to Face Recognition

In this section, we present an application of $GiFME$ to face recognition. We use the *YaleExtendedB* [20] face recognition database to conduct our experiments. This database contains 16128 face images taken from 28 human subjects with varying pose and illumination conditions. In order to do the recognition, we formulate the face recognition problem as one that requires the computation of the intrinsic mean on the Grassmann manifold, as described in the following paragraph.

For each person indexed by i, consider his/her face images. From each image, j, construct the SIFT descriptor matrix, $\mathcal{F}^{ij}$ [22]. Let the dimension of each $\mathcal{F}^{ij}$ be $n \times m$, $n > m$. For each (i, j), construct the matrix P^{ij} where the kth column of the matrix is the kth principal vector of $\mathcal{F}^{ij}$. Suppose, we have considered the first $c, 1 \leq c < m$ principal vectors. Then, we can view each of $\mathcal{F}^{ij}$ as a point on $Grass(c, m)$ such that each row of $\mathcal{F}^{ij}$ belongs to the c dimensional subspace of $\mathbf{R}^m$. Then, for i^{th} person, construct the geodesic submanifold, H_i of $Grass(c, m)$ according to [11]. In order to construct the submanifold, we first compute the FM (on the Grassmann) μ_i of $\mathcal{F}^{ij}$, $\forall j$. Then get the first c principal components of $Exp_{\mu_i}^{-1}\mathcal{F}^{ij}$, $\forall j$, where Exp^{-1} is the Riemannian Inverse Exponential

Method	Precision(%)	Time(s)
GFME	**98.77**	6808.8
GiFME	98.40	**16.31**

Table 1. Comparison results on *Yale* database

Method	Scenario	Precision(%)	Time(s)
GFME	d1	**83.24**	2714.71
GiFME	d1	81.36	**6.55**
GFME	d2	**80.64**	2763.91
GiFME	d2	80.00	**6.92**
GFME	d3	**87.89**	4260.67
GiFME	d3	80.68	**17.54**
GFME	d4	**95.92**	3973.83
GiFME	d4	91.84	**17.73**

Table 2. Comparison results on the *KTH* action recognition database

Map. These c principal vectors form the orthonormal basis of $T_{\mu_i}H_i$ where, H_i is the geodesic submanifold. So, for each person i, we construct the orthonormal basis of $T_{\mu_i}H_i$. Then, given an unknown face image, we first extract from it, the SIFT descriptor matrix, say $\mathcal{G} \in Grass(c,m)$. Define the projection operator, $\pi_{H_i} : Grass(c,m) \rightarrow H_i$ by $\pi_{H_i}(\mathcal{G}) = \arg \min_{\mathcal{X} \in H_i} d_{Gr}(\mathcal{X}, \mathcal{G})$ As discussed in [11], projection onto a geodesic submanifold can be approximated linearly in the tangent space. First project G onto $T_{\mu_i}H_i$ by using Riemannian Inverse Exponential map, let $v_{\mathcal{G}} = Exp_{\mu_i}^{-1}\mathcal{G}$. Let $\{v_j^i\}$ be the orthonormal basis of $T_{mu_i}H_i$. Then, $\pi_{H_i}(\mathcal{G})$ can be approximated as follows:

$$\pi_{H_i}(\mathcal{G}) = Exp_{\mu_i}\left(\sum_j \langle v_{\mathcal{G}}, v_j^i \rangle \ v_j^i \right)$$ where, Exp is the

Riemannian Exponential map, $\langle .,* \rangle$ is the inner product, which can be taken as the standard dot product since an arbitrary inner product can be taken as dot product if there is an orthonormal basis. Then, we assign $\mathcal{G}$ to class $\tilde{c}$, if

$$\tilde{c} = \arg \min_j d_{Gr}\left(\pi_{H_j}(\mathcal{G}), \mu_j \right)$$

For each class (person), we use $90\% - 10\%$ partition and use 90% to compute the orthonormal basis of $T_{\mu_i}H_i$ and use the remaining 10% to predict the class. We have compared the performance of $GiFME$ and $GFME$ both in terms of precision and running time. Note that our main emphasis here is on the efficient computation of the FM, rather than improving the precision over the state-of-the-art face recognition systems. The comparative results in terms of computation time and precision are given in Table 1. From the table, it is evident that though the precision for both $GFME$ and $GiFME$ are comparable, $GiFME$ outperforms the $GFME$ in the running time to compute the FM.

3.3. Application to Video Action Recognition

In this section, we present an application of the FM computation on the Grassmann to video face recognition. We have used the *KTH Action* database [28] which contains 6 actions performed by 25 human subjects in 4 scenarios (denoted by 'd1', 'd2', 'd3', 'd4'). All videos are captured by static camera with homogeneous background.

For each video, we first extract 25 frames, then extract the HOG features [8] from each of the frame. Then, each video will become a point on $Grass(p,n)$ with appropriate p and n values, similar to the Face recognition application.

Then, we use the same algorithm as before to compute the orthonormal basis for each action class. Given a video of an unknown action class, we first get the corresponding point on $Grass(p,n)$ after extracting HOG features from each frame. Then, assign this point to the class for which the projection error is minimum. We have performed leave one out experiments to get the precision of recognition. Similar to earlier experiments, we compare the performance of $GiFME$ over $GFME$ both for precision as well as time for computing the FM. As evident from results presented in Table 2, we can see that the precision using $GiFME$ is slightly less than that of $GFME$ but in terms of computation time, $GiFME$ is significantly better.

4. Conclusions

In this paper, we have addressed the problem of computing the FM of data residing on a Grassmann manifold. This is a commonly encountered problem in Computer Vision with numerous applications. Computing the FM is a computationally intensive task and most existing techniques resort to a gradient descent approach in minimizing the sum of squared geodesic distances on the Grassmann. We presented an efficient recursive algorithm for estimating the FM namely, GiFME. The key contributions of this paper are, (i) a proof of convergence (in probability) of GiFME to the FM of the underlying distribution. This is basically the weak consistency of GiFME. (ii) We demonstrated through experiments, the significant gain in compute time over the batch mode counterpart (GFME). (iii) Finally, we show the performance of GiFME on several real datasets drawn from the domain of face and action recognition. Our future work will involve extending the inductive FM estimators to the close relative of Grassmann namely, the Stiefel manifold.

Acknowledgement

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80(2):199–220, 2004.

[2] B. Afsari. Riemannian L^{p} center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011.

[3] B. Afsari, R. Tron, and R. Vidal. On the convergence of gradient descent for finding the riemannian center of mass. *SIAM Journal on Control and Optimization*, 51(3):2230–2260, 2013.

[4] G. Aggarwal, A. K. R. Chowdhury, and R. Chellappa. A system identification approach for video-based face recognition. In *ICPR*, volume 4, pages 175–178, 2004.

[5] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto. Recognition of human gaits. In *CVPR*, volume 2, pages II–52, 2001.

[6] H. E. Cetingul and R. Vidal. Intrinsic mean shift for clustering on stiefel and grassmann manifolds. In *CVPR*, pages 1896–1902, 2009.

[7] Y. Chikuse. *Statistics on Special Manifolds*. Springer, February 2003.

[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893, 2005.

[9] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIMAX*, 20(2):303–353, 1998.

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

[11] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE TMI*, 23(8):995–1005, 2004.

[12] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré*, volume 10, pages 215–310. Presses universitaires de France, 1948.

[13] C. R. Goodall and K. V. Mardia. Projective shape analysis. *Journal of Computational and Graphical Statistics*, 8(2):143–168, 1999.

[14] D. Groisser. Newton's method, zeroes of vector fields, and the riemannian center of mass. *Advances in Applied Mathematics*, 33(1):95–135, 2004.

[15] S. Hauberg, A. Feragen, and M. J. Black. Grassmann averages for scalable robust PCA. In *CVPR*, pages 3810 –3817, 2014.

[16] J. Ho, G. Cheng, H. Salehian, and B. Vemuri. Recursive Karcher expectation estimators and geometric law of large numbers. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pages 325–332, 2013.

[17] H. Hopf and W. Rinow. Über den begriff der vollständigen differentialgeometrischen fläche. *Commentarii Mathematici Helvetici*, 3(1):209–225, 1931.

[18] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5):509–541, 1977.

[19] S. Kobayashi and K. Nomizu. *Foundations of differential geometry*. New York, 1963.

[20] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE TPAMI*, 27(5):684–698, 2005.

[21] Y. Lim and M. Pálfia. Weighted inductive means. *Linear Algebra and its Applications*, 453:59–83, 2014.

[22] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[23] Y. M. Lui. Advances in matrix manifolds for computer vision. *Image and Vision Computing*, 30(6):380–388, 2012.

[24] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIMAX*, 26(3):735–747, 2005.

[25] V. Patrangenaru and K. V. Mardia. Affine shape analysis and image analysis. In *22nd Leeds Annual Statistics Research Workshop*, 2003.

[26] H. Salehian, R. Chakraborty, E. Ofori, D. Vaillancourt, and B. C. Vemuri. An efficient recursive estimator of the fréchet mean on a hypersphere with applications to medical image analysis. *Mathematical Foundations of Computational Anatomy*, 2015.

[27] H. Salehian, G. Cheng, B. C. Vemuri, and J. Ho. Recursive estimation of the Stein center of SPD matrices and its applications. In *ICCV*, pages 1793–1800. IEEE, 2013.

[28] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: a local svm approach. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 3, pages 32–36, 2004.

[29] A. Schwartzman. *Random ellipsoids and false discovery rates: Statistics for diffusion tensor imaging data*. PhD thesis, Stanford University, 2006.

[30] A. Srivastava and E. Klassen. Bayesian and geometric subspace tracking. *Advances in Applied Probability*, pages 43–56, 2004.

[31] K.-T. Sturm. Probability measures on metric spaces of non-positive curvature. *Contemporary mathematics*, 338:357–390, 2003.

[32] R. Subbarao and P. Meer. Discontinuity preserving filtering over analytic manifolds. In *CVPR*, pages 1–6, 2007.

[33] P. Turaga and R. Chellappa. Nearest-neighbor search algorithms on non-euclidean manifolds for computer vision applications. In *ICVGIP*, pages 282–289, 2010.

[34] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa. Statistical computations on grassmann and stiefel manifolds for image and video-based recognition. *IEEE TPAMI*, 33(11):2273–2286, 2011.

[35] A. Veeraraghavan, A. K. Roy-Chowdhury, and R. Chellappa. Matching shape sequences in video with applications in human movement analysis. *IEEE TPAMI*, 27(12):1896–1909, 2005.