*The Annals of Statistics*
2019, Vol. 47, No. 1, 415–438
https://doi.org/10.1214/18-AOS1692


# STATISTICS ON THE STIEFEL MANIFOLD: THEORY AND APPLICATIONS[1]

BY RUDRASIS CHAKRABORTY AND BABA C. VEMURI

*University of Florida*

A Stiefel manifold of the compact type is often encountered in many fields of engineering including, signal and image processing, machine learning, numerical optimization and others. The Stiefel manifold is a Riemannian homogeneous space but not a symmetric space. In previous work, researchers have defined probability distributions on symmetric spaces and performed statistical analysis of data residing in these spaces. In this paper, we present original work involving definition of Gaussian distributions on a homogeneous space and show that the maximum-likelihood estimate of the location parameter of a Gaussian distribution on the homogeneous space yields the Fréchet mean (FM) of the samples drawn from this distribution. Further, we present an algorithm to sample from the Gaussian distribution on the Stiefel manifold and recursively compute the FM of these samples. We also prove the weak consistency of this recursive FM estimator. Several synthetic and real data experiments are then presented, demonstrating the superior computational performance of this estimator over the gradient descent based nonrecursive counter part as well as the stochastic gradient descent based method prevalent in literature.

**1. Introduction.** Manifold-valued data have gained much importance in recent times due to their expressiveness and ready availability of machines with powerful CPUs and large storage. For example, these data arise as *rank-2 tensors* (manifold of symmetric positive definite matrices) [Moakher (2006), Pennec, Fillard and Ayache (2006)], *linear subspaces* (the Grassmann manifold) [Goodall and Mardia (1999), Hauberg, Feragen and Black (2014), Patrangenaru and Mardia (2003), Turaga, Veeraraghavan and Chellappa (2008)], *column orthogonal matrices* (the Stiefel manifold) [Chikuse (1991), Hendriks and Landsman (1998), Turaga, Veeraraghavan and Chellappa (2008)], *directional data* and *probability densities* (the hypersphere) [Hartley et al. (2013), Mardia and Jupp (2000), Srivastava, Jermyn and Joshi (2007), Tuch et al. (2003) and others]. A useful method of analyzing manifold valued data is to compute statistics on the underlying manifold. The most popular statistic is a *summary* of the data, that is, *the* Rie-

Received February 2017; revised February 2018.
[1]Supported in part by the NSF Grants IIS-1525431 and IIS-1724174.
*MSC2010 subject classifications.* Primary 62F12; secondary 58A99.
*Key words and phrases.* Homogeneous space, Stiefel manifold, Fréchet mean, Gaussian distribution.

mannian barycenter [Fréchet mean (FM)] [Afsari (2011), Fréchet (1948), Karcher (1977)], Fréchet median [Arnaudon, Barbaresco and Yang (2013), Charfi et al. (2013), etc.]. However, in order to compute statistics of manifold-valued data, the first step involves defining a distribution on the manifold. Recently, authors in Said et al. (2016) have defined a Gaussian distribution on Riemannian symmetric spaces (or symmetric spaces). Some typical examples of symmetric spaces include the Grassmannian, the hypersphere, etc. Several other researchers [Cheng and Vemuri (2013), Said et al. (2017)] have defined a Gaussian distribution on the space of symmetric positive definite matrices. They called the distribution a "generalized Gaussian distribution" [Cheng and Vemuri (2013)] and "Riemannian Gaussian distribution" [Said et al. (2017)], respectively.

In this work, we define a Gaussian distribution on a homogeneous space (a more general class than symmetric spaces). A key difficulty in defining the Gaussian distribution on a non-Euclidean space is to show that the normalizing factor in the expression for the distribution is a constant. In this work, we show that the normalizing factor in our definition of the Gaussian distribution on a homogeneous space is indeed a constant. Note that a symmetric space is a homogeneous space but not all homogeneous spaces are symmetric, and thus, our definition of Gaussian distribution is on a more generalized topological space than the symmetric space. Given a well-defined Gaussian distribution, the next step is to estimate the parameters of the distribution. In this work, we prove that the maximum likelihood estimate (MLE) of the mean of the Gaussian distribution is the Fréchet mean (FM) of the samples drawn from the distribution.

Data with values in the space of column orthogonal matrices have become popular in many applications of Computer Vision and Medical Image analysis [Cetingul and Vidal (2009), Chakraborty, Banerjee and Vemuri (2017), Lui (2012), Pham and Venkatesh (2008), Turaga, Veeraraghavan and Chellappa (2008)]. The space of column orthogonal matrices is a topological space, and moreover, one can equip this space with a Riemannian metric which in turn makes this space a Riemannian manifold, known as the Stiefel manifold. The Stiefel manifold is a homogeneous space and here we extend the definition of the Gaussian distribution to the Stiefel manifold. In this work, we restrict ourselves to the Stiefel manifold of the compact type, which is quite commonly encountered in most applications mentioned earlier.

We now motivate the need for a recursive FM estimator. In this age of massive and continuous streaming data, samples are often acquired incrementally. Hence, from an applications perspective, the desired algorithm should be *recursive/inductive* in order to maximize computational efficiency and account for availability of data, requirements that are seldom addressed in more theoretically oriented fields. We propose an *inductive* FM computation algorithm and prove the weak consistency of our proposed estimator. FM computation on Riemannian manifolds has been an active area of research for the past few decades. Several researchers have addressed this problem and we refer the reader to Afsari (2011), Ando, Li and Mathias (2004), Arnaudon, Barbaresco and Yang (2013), Bhatia (1997), Bhattacharya and Bhattacharya (2008), Chakraborty and Vemuri (2015), Fletcher and Joshi (2007), Groisser (2004), Ho et al. (2013), Moakher (2005), Pennec (2006), Rao (1987), Salehian et al. (2015), Sturm (2003).

1.1. *Key contributions*. In summary, the key contributions of this paper are: (i) A novel generalization of Gaussian distributions to compact homogeneous spaces. (ii) A proof that the MLE of the location parameter of this distribution is "the" FM. (iii) A sampling technique for drawing samples from this generalized Gaussian distribution defined on a compact Stiefel manifold (which is a homogeneous space), and an inductive/recursive FM estimator from the drawn samples along with a proof of its weak consistency. Several examples of FM estimates computed from real and synthetic data are shown to illustrate the power of the proposed methods.

Though researchers have defined Gaussian distributions on other manifolds in the past [see Cheng and Vemuri (2013), Said et al. (2016)], their generalization of the Gaussian distribution is restricted to symmetric spaces of noncompact types. In this work, we define a Gaussian distribution on a compact homogeneous space, which is a more general topological space than the symmetric space. A few others in literature have generalized the Gaussian distribution to all Riemannian manifolds, for instance, in Zhang and Fletcher (2013), authors defined the Gaussian distribution on a Riemannian manifold without a proof to show that the normalizing factor is a constant. In Grenander (2008), author proposed a generalized Gaussian distribution as a solution to the heat equation. In Fletcher (2013), though the author commented on the constancy of the normalizing factor for Riemannian homogeneous spaces, he did not however prove the finiteness of the normalization factor. It should be noted that the finiteness of the normalization factor is crucial for the proposed distribution to be a valid distribution. In Pennec (2006), the author defined the normal law on Riemannian manifolds using the concept of entropy maximization for distributions with known mean and covariance. Under certain assumptions, the author shows that this definition amounts to using the Riemannian exponential map on a truncated Gaussian distribution defined in the tangent space at the known intrinsic mean. This approach of deriving the normal distribution yields a normalizing factor that is dependent on the location parameter of the distribution, and hence is not a constant with respect to the FM.

We then move our focus to the Stiefel manifold (which is a homogeneous space) and propose a simple algorithm to draw samples from the Gaussian distribution on the Stiefel manifold. In order to achieve this, we develop a simple but nontrivial way to extend the sampling algorithm in Said et al. (2016) to get samples on the Stiefel manifold. Once we have the samples from a Gaussian distribution on the Stiefel, we propose a novel estimator of the sample FM and prove the weak consistency of this estimator. The proposed FM estimator is inductive in nature and is motivated by the inductive FM algorithm on the Euclidean space. But, unlike Euclidean space, due to the presence of nonzero curvature, it is necessary to prove the consistency of our proposed estimator, which is presented subsequently. Further, we experimentally validate the superior performance of our proposed FM estimator over the gradient descent based techniques. Moreover, we also show that the MLE of the location parameter of the Gaussian distribution on the Stiefel manifold asymptotically achieves the Cramér–Rao lower bound [Cramér (1946), Rao (1992)], hence in turn, the MLE of the location parameter is efficient. This implies that our proposed consistent FM estimator, asymptotically, has a variance lower bounded by that of the MLE.

The rest of the paper is organized as follows. In Section 2, we present the necessary mathematical background. In Section 3, we define a Gaussian distribution on a homogeneous space. More specifically, define a generalized Gaussian distribution on the Stiefel manifold and prove that the normalizing factor is indeed a constant with respect to the location parameter of the distribution. Then we propose a sampling algorithm to draw samples from this generalized Gaussian distribution in Section 3.1 and in Section 3.2, show that the MLE of the location parameter of this Gaussian distribution is the FM of the samples drawn from the distribution. In Section 4, we propose an inductive FM estimator and prove its weak consistency. Finally, we present a set of synthetic and real data experiments in Section 5 and draw conclusions in Section 6.

**2. Mathematical background: Homogeneous spaces and the Riemannian symmetric space.** In this section, we present a brief note on the differential geometry background required in the rest of the paper. For a detailed exposition on these concepts, we refer the reader to a comprehensive and excellent treatise on this topic by Helgason [Helgason (1978)]. Several propositions and lemmas that are needed to prove the results in the rest of the paper are stated and proved here. Some of these might have been presented in the vast differential geometry literature but are unknown to us, and hence the proofs presented in this background section are original.

Let $(\mathcal{M}, g^{\mathcal{M}})$ be a Riemannian manifold with a Riemannian metric $g^{\mathcal{M}}$, that is, $(\forall x \in \mathcal{M})$ $g_x^{\mathcal{M}} : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbf{R}$ is a bi-linear symmetric positive definite map, where $T_x\mathcal{M}$ is the tangent space of $\mathcal{M}$ at $x \in \mathcal{M}$. Let $d : \mathcal{M} \times \mathcal{M} \to \mathbf{R}$ be the metric (distance) induced by the Riemannian metric $g^{\mathcal{M}}$. Let $I(\mathcal{M})$ be the set of all isometries of $\mathcal{M}$, that is, given $g \in I(\mathcal{M})$, $d(g.x, g.y) = d(x, y)$, for all $x, y \in \mathcal{M}$. It is clear that $I(\mathcal{M})$ forms a group [henceforth, we will denote $I(\mathcal{M})$ by $(G, \cdot)$], and thus, for a given $g \in G$ and $x \in \mathcal{M}$, $g.x \mapsto y$, for some $y \in \mathcal{M}$ is a group action. Consider $o \in \mathcal{M}$, and let $H = \mathrm{Stab}(o) = \{h \in G | h.o = o\}$, that is, $H$ is the *Stabilizer* of $o \in \mathcal{M}$. We say that $G$ acts *transitively* on $\mathcal{M}$, *iff*, given $x, y \in \mathcal{M}$, there exists a $g \in \mathcal{M}$ such that $y = g.x$.

DEFINITION 2.1. Let $G = I(\mathcal{M})$ act transitively on $\mathcal{M}$ and $H = \mathrm{Stab}(o)$, $o \in \mathcal{M}$ (called the "origin" of $\mathcal{M}$) be a subgroup of $G$. Then $\mathcal{M}$ is a homogeneous space and can be identified with the quotient space $G/H$ under the diffeomorphic mapping $gH \mapsto g.o$, $g \in G$ [Helgason (1978)].

In fact, if $\mathcal{M}$ is a homogeneous space, then $G$ is a Lie group. A Stiefel manifold, $\mathrm{St}(p, n)$ (definition of the Stiefel manifold is given in next section) is a homogeneous space and can be identified with $O(n)/O(n-p)$, where $O(n)$ is the group of orthogonal matrices. Now, we will list some of the important properties of homogeneous spaces that will be used throughout the rest of the paper.

*Properties of homogeneous spaces*: Let $(\mathcal{M}, g^{\mathcal{M}})$ be a homogeneous space. Let $\omega^{\mathcal{M}}$ be the corresponding volume form and $F : \mathcal{M} \to \mathbf{R}$ be any integrable function. Let $g \in G$, s.t. $y = g.x$, $x, y \in \mathcal{M}$. Then the following facts are true:

1. $g^{\mathcal{M}}(dy, dy) = g^{\mathcal{M}}(dx, dx)$.
2. $d(x, z) = d(y, g.z)$, for all $z \in \mathcal{M}$.
3. $\int_{\mathcal{M}} F(y)\omega^{\mathcal{M}}(x) = \int_{\mathcal{M}} F(x)\omega^{\mathcal{M}}(x)$.

DEFINITION 2.2. A Riemannian symmetric space is a Riemannian manifold $\mathcal{M}$ with the following property: $(\forall x \in \mathcal{M})(\exists s_x \in G)$ such that $s_x.x = x$ and $ds_x|_x = -I$. $s_x$ is called *symmetry* at $x$ [Helgason (1978)].

PROPOSITION 2.1 [Helgason (1978)]. *A symmetric space $\mathcal{M}$ is a homogeneous space with a symmetry, $s_o$, at $o \in \mathcal{M}$. For the other point $x \in \mathcal{M}$, by transitivity of $G$, there exists $g \in G$ such that $x = g.o$ and $s_x = g \cdot s_o \cdot g^{-1}$.*

PROPOSITION 2.2 [Helgason (1978)]. *Any symmetric space is geodesically complete.*

Some examples of symmetric spaces include, $\mathbf{S}^n$ (the hypersphere), $\mathbf{H}^n$ (the hyperbolic space) and $\mathrm{Gr}(p, n)$ (the Grassmannian). It is evident from the definition that symmetric space is a homogeneous space but the converse is not true. For example, the Stiefel manifold is not a symmetric space.

PROPOSITION 2.3 [Helgason (1978)]. *The mapping $\sigma : g \mapsto s_o \cdot g \cdot s_o$ is an involutive automorphism of $G$ and the stabilizer of $o$, that is, $H$, is contained in the group of fixed points of $\sigma$.*

Clearly, $\sigma(e) = e$, as $\sigma$ is an automorphism, $e \in G$ is the identity element. Recall, $G$ is a Lie group, hence, differentiating $\sigma$ at $e$, we get an involutive automorphism of the Lie algebra $\mathfrak{g}$ of $G$ (also denoted by $\sigma$). Henceforth, we will use $\sigma$ to denote the automorphism of $\mathfrak{g}$. Since $\sigma$ is involutive, that is, $\sigma^2 = I$, $\sigma$ has two eigenvalues, $\pm 1$ and let $\mathfrak{h}$ (Lie algebra of $H$) and $\mathfrak{p}$ be the corresponding eigenspaces, then $\mathfrak{g} = \mathfrak{h} + \mathfrak{p}$ (direct sum).

PROPOSITION 2.4 [Helgason (1978)]. $[\mathfrak{h}, \mathfrak{h}] \subseteq \mathfrak{h}$, $[\mathfrak{h}, \mathfrak{p}] \subseteq \mathfrak{p}$ *and* $[\mathfrak{p}, \mathfrak{p}] \subseteq \mathfrak{h}$.

Hence, $\mathfrak{h}$ is a Lie subalgebra of $\mathfrak{g}$. Henceforth, we will assume $\mathfrak{g}$ to be semisimple. We can define a symmetric, bilinear form, $B$ on $\mathfrak{g}$ as follows: $B(u, v) = \mathrm{trace}(\mathrm{ad}(u) \circ \mathrm{ad}(v))$, where $\mathrm{ad}(u)$ is the adjoint endomorphism of $\mathfrak{g}$ defined by $\mathrm{ad}(u)(v) = [u, v]$. $B$ is called the *Killing form* on $\mathfrak{g}$.

DEFINITION 2.3. The decomposition of $\mathfrak{g}$ as $\mathfrak{g} = \mathfrak{h} + \mathfrak{p}$ is called the Cartan decomposition of $\mathfrak{g}$ associated with the involution $\sigma$. Furthermore, $B$ is negative definite on $\mathfrak{h}$, positive definite on $\mathfrak{p}$ and $\mathfrak{h}$ and $\mathfrak{p}$ are an orthogonal complement of each other with respect to $B$ on $\mathfrak{g}$.

Recall, a symmetric space, $\mathcal{M}$, can be identified with $G/H$. Note that, $o$, the "origin" of $\mathcal{M}$ can be written as $o = eH$, $e \in G$ is the identity element. Since, $\mathfrak{p}$ can be identified with $T_o\mathcal{M}$, the Riemannian metric $g^{\mathcal{M}}$ on $\mathcal{M}$ corresponds to the Killing form $B$ on $\mathfrak{p}$ [Helgason (1978)], which is a $H$-invariant form. Without loss of generality, we will assume that $\mathfrak{g}$ is over $\mathbf{R}$ and $\mathfrak{g}$ be semisimple (equivalently, the Killing form on $\mathfrak{g}$ is nondegenerate). The symmetric space $G/H$ is said to be *compact* (*noncompact*) *iff* the sectional curvature is strictly positive (negative), equivalently *iff* $\mathfrak{g}$ is compact (noncompact).

*Duality*: Given a semisimple Lie algebra $\mathfrak{g}$ with the Cartan decomposition $\mathfrak{g} = \mathfrak{h} + \mathfrak{p}$, construct another Lie algebra $\tilde{\mathfrak{g}}$ from $\mathfrak{g}$ as follows: $\tilde{\mathfrak{g}} = \mathfrak{h} + J(\mathfrak{p})$, where $J$ is a complex structure of $\mathfrak{p}$ (real Lie algebra). From the definition of complex structure, $J : \mathfrak{p} \to \mathfrak{p}$, is an automorphism on $\mathfrak{p}$ s.t., $J^2 = -I$. $J$ satisfies the following equality: $J([T, W]) = [J(T), W] = [T, J(W)]$, for all $T, W \in \mathfrak{p}$. We will call $\tilde{\mathfrak{g}}$ the dual Lie algebra of $\mathfrak{g}$. It is easy to see that if $\mathfrak{g}$ corresponds to a symmetric space of noncompact type, $\tilde{\mathfrak{g}}$ is a symmetric space of compact type and vice versa. *This duality property is very useful and is a key ingredient of this paper*.

Now, we will briefly describe the geometry of two Riemannian manifolds, namely the Stiefel manifold and the Grassmannian. We need the geometry of the Stiefel manifold throughout the rest of the paper. Furthermore, observe that the Stiefel and the Grassmannian form a fiber bundle. In order to draw samples from a distribution on the Stiefel, we will use the samples drawn from a distribution on the Grassmannian by exploiting the fiber bundle structure. Hence, we will require the geometry of the Grassmannian as well, which we will briefly present below.

*Differential geometry of the Stiefel manifold*: The set of all full column rank $(n \times p)$ dimensional real matrices form a Stiefel manifold, $\mathrm{St}(p, n)$, where $n \geq p$. A compact Stiefel manifold is the set of all column orthonormal real matrices. When $p < n$, $\mathrm{St}(p, n)$ can be identified with $\mathrm{SO}(n)/\mathrm{SO}(n-p)$, where $\mathrm{SO}(m)$ is $m \times m$ special orthogonal group. Note that, when we consider the quotient space, $\mathrm{SO}(n)/\mathrm{SO}(n-p)$, we assume that $\mathrm{SO}(n-p) \simeq F(\mathrm{SO}(n-p))$ is a subgroup of $\mathrm{SO}(n)$, where $F : \mathrm{SO}(n-p) \to \mathrm{SO}(n)$ defined by $X \mapsto \left[\begin{smallmatrix} I_p & 0 \\ 0 & X \end{smallmatrix}\right]$ is an isomorphism from $\mathrm{SO}(n-p)$ to $F(\mathrm{SO}(n-p))$.

PROPOSITION 2.5. $\mathrm{SO}(n-p)$ *is a closed Lie-subgroup of* $\mathrm{SO}(n)$. *Moreover, the quotient space* $\mathrm{SO}(n)/\mathrm{SO}(n-p)$ *together with the projection map,* $\Pi : \mathrm{SO}(n) \to \mathrm{SO}(n)/\mathrm{SO}(n-p)$ *is a principal bundle with* $\mathrm{SO}(n-p)$ *as the fiber.*

PROOF. $\mathrm{SO}(n-p)$ is a compact Lie-subgroup of $\mathrm{SO}(n)$, hence $\mathrm{SO}(n-p)$ is a closed subgroup. The fiber bundle structure of $(\mathrm{SO}(n), \mathrm{SO}(n)/\mathrm{SO}(n-p), \Pi)$ follows directly from the closedness of $\mathrm{SO}(n-p)$. As $\mathrm{SO}(n)$ is a principal homogeneous space [because $\mathrm{SO}(n) \simeq \mathrm{St}(n-1, n)$ and $\mathrm{SO}(n)$ acts on it freely], hence the principal bundle structure. □

With a slight abuse of notation, henceforth, we denote the compact Stiefel manifold by $\mathrm{St}(p, n)$. Hence, $\mathrm{St}(p, n) = \{X \in \mathbf{R}^{n\times p} | X^T X = I_p\}$, where $I_p$ is the $p \times p$ identity matrix. The compact Stiefel manifold has dimension $pn - \frac{p(p+1)}{2}$. At any $X \in \mathrm{St}(p, n)$, the tangent space $T_X \mathrm{St}(p, n)$ is defined as follows: $T_X \mathrm{St}(p, n) = \{U \in \mathbf{R}^{n\times p} | X^T U + U^T X = 0\}$. Now, given $U, V \in T_X \mathrm{St}(p, n)$, the canonical Riemannian metric on $\mathrm{St}(p, n)$ is defined as follows:

$$\langle U, V\rangle_X = \mathrm{trace}(U^T V). \tag{2.1}$$

With this metric, the compact Stiefel manifold has nonnegative sectional curvature [Ziller (2007)].

Given $X \in \mathrm{St}(p, n)$, we can define the Riemannian retraction and lifting map within an open neighborhood of $X$. We will use an efficient Cayley-type retraction and lifting maps, respectively, on $\mathrm{St}(p, n)$ as defined in Fraikin, Hüper and Dooren (2007), Kaneko, Fiori and Tanaka (2013). It should be mentioned that though the domain of retraction is a subset of the domain of inverse-Exponential map, on $\mathrm{St}(p, n)$ retraction/ lifting is a useful alternative since *there are no closed-form expressions for both the Exponential and the inverse-Exponential maps on the Stiefel manifold*. Recently, a fast iterative algorithm to compute Riemannian inverse-Exponential map has been proposed in Zimmermann (2017), which can be used instead of retraction/lifting maps to compute FM in our algorithm.

In the neighborhood of $[I_p\ 0]$ ($n \times p$ matrix with upper-right $p \times p$ block is identity and rest are zeros), given $X \in \mathrm{St}(p, n)$, we define the lifting map $\mathrm{Exp}_X^{-1} : \mathrm{St}(p, n) \to T_X \mathrm{St}(p, n)$ by $\mathrm{Exp}_X^{-1}(Y) = \left[\begin{smallmatrix} C & -B^T \\ B & 0 \end{smallmatrix}\right]$ where, $C$ is a $p \times p$ skew-symmetric matrix and $B$ is a $(n-p) \times p$ matrix defined as follows: $C = 2(X_u^T + Y_u^T)^{-1} \mathrm{sk}(Y_u^T X_u + X_l^T Y_l)(X_u + Y_u)^{-1}$ and $B = (Y_l - X_l)(X_u + Y_u)^{-1}$ where, $X = [X_u, X_l]^T$, and $Y = [Y_u, Y_l]^T$ with $X_u, Y_u \in \mathbf{R}^{p\times p}$, and $X_l, Y_l \in \mathbf{R}^{(n-p)\times p}$, provided that $X_u + Y_u$ is nonsingular. $\mathrm{sk}(M)$ is defined as $\frac{1}{2}(M^T - M)$ and, $Y \in \mathrm{St}(p, n)$.

Furthermore, in the neighborhood of $[I_p\ 0]$, the retraction map defined above is a diffeomorphism (since it is a chart map) from $\mathrm{St}(p, n)$ to $\mathfrak{so}(n)$.

PROPOSITION 2.6. *The projection map* $\Pi : \mathrm{SO}(n) \to \mathrm{SO}(n)/\mathrm{SO}(n-p)$ *is a covering map on the neighborhood of* $\mathrm{SO}(n-p)$ *in* $\mathrm{SO}(n)/\mathrm{SO}(n-p)$.

PROOF. First, note that under the identification of $\mathrm{St}(p,n)$ with $\mathrm{SO}(n)/\mathrm{SO}(n-p)$, the neighborhood of $[I_p\ 0]$ in $\mathrm{St}(p,n)$ can be identified with the neighborhood of $\mathrm{SO}(n-p)$ in $\mathrm{SO}(n)/\mathrm{SO}(n-p)$. Now, the retraction map defined above is a (local) diffeomorphism from $\mathrm{St}(p,n)$ to $\mathfrak{so}(n)$. Also, the Cayley map is a diffeomorphism from $\mathfrak{so}(n)$ to the neighborhood of $I_n$ in $\mathrm{SO}(n)$. Thus, the map $\Pi : \mathrm{SO}(n) \to \mathrm{SO}(n)/\mathrm{SO}(n-p)$ is a diffeomorphism to the neighborhood of $\mathrm{SO}(n-p)$ in $\mathrm{SO}(n)/\mathrm{SO}(n-p)$ (using the fact that the composition of two diffeomorphisms is a diffeomorphism). Now, since $\mathrm{SO}(n)$ is compact and $\Pi$ is surjective, $\Pi$ is a covering map on the neighborhood of $\mathrm{SO}(n-p)$ in $\mathrm{SO}(n)/\mathrm{SO}(n-p)$ using the following lemma. □

LEMMA 2.1. *Under the hypothesis in Proposition* 2.6, $\Pi : \mathrm{SO}(n) \to \mathrm{SO}(n)/\mathrm{SO}(n-p)$ *is a covering map in the neighborhood from the neighborhood of* $I_n$ *in* $\mathrm{SO}(n)$ *to the neighborhood of* $\mathrm{SO}(n-p)$ *of* $\mathrm{SO}(n)/\mathrm{SO}(n-p)$.

PROOF. In Proposition 2.6, we have shown that $\Pi$ is local diffeomorphism in the neighborhood specified in the hypothesis. Let $\mathcal{V}$ be a neighborhood around $\mathrm{SO}(n-p)$ in $\mathrm{SO}(n)/\mathrm{SO}(n-p)$ and $\mathcal{U}$ be a neighborhood around $I_n$ in $\mathrm{SO}(n)$ on which $\Pi$ is a diffeomorphism. Let $Y \in \mathcal{V}$, as $\mathrm{SO}(n)/\mathrm{SO}(n-p)$ is a $T_2$ space, hence, $\{Y\}$ is closed, thus $\Pi^{-1}(Y)$ is closed and since $\mathrm{SO}(n)$ is compact, hence $\Pi^{-1}(Y)$ is compact. For each $X \in \Pi^{-1}(Y)$, let $\mathcal{U}_X$ be a open neighborhood around $X$ where $\Pi$ restricts to a diffeomorphism (and hence homeomorphism). Then $\{\mathcal{U}_X : X \in \Pi^{-1}(Y)\}$ is an open cover of $\Pi^{-1}(Y)$, thus as a finite subcover $\{\mathcal{U}_X\}_{X \in I}$, where $I$ is finite. We chose $\{\mathcal{U}_X\}$ to be disjoint as $\mathrm{SO}(n)$ is a $T_2$ space. Let $\mathcal{W} = \bigcap_{X \in I} \Pi(\mathcal{U}_X)$, which is an open neighborhood of $Y$. Then $\{\Pi^{-1}(\mathcal{W}) \cap \mathcal{U}_X\}_{X \in I}$ is a disjoint collection of open neighborhoods each of which maps homeomorphically to $\mathcal{V}$. Hence, $\Pi$ is a covering map in the local neighborhood. □

Given $W \in \mathfrak{so}(n)$, the Cayley map is a conformal mapping, $\mathrm{Gr} : \mathfrak{so}(n) \to \mathrm{SO}(n)$ defined by $\mathrm{Gr}(W) = (I_n + W)(I_n - W)^{-1}$. Using the Cayley mapping, we can define the Riemannian retraction map $\mathrm{Exp}_X : T_X \mathrm{St}(p,n) \to \mathrm{St}(p,n)$ by $\mathrm{Exp}_X(W) = \mathrm{Gr}(W)X$. Hence, given $X, Y \in \mathrm{St}(p,n)$ within a regular geodesic ball (the geodesic ball does not include the cut locus) of appropriate radius (henceforth, we will assume the geodesic ball to be regular), we can define the unique geodesic from $X$ to $Y$, denoted by $\Gamma_X^Y(t)$ as

$$\Gamma_X^Y(t) = \mathrm{Exp}_X\big(t\,\mathrm{Exp}_X^{-1}(Y)\big). \tag{2.2}$$

Also, we can define the distance between $X$ and $Y$ as

$$d(X,Y) = \sqrt{\langle \mathrm{Exp}_X^{-1}(Y), \mathrm{Exp}_X^{-1}(Y)\rangle}. \tag{2.3}$$

*Differential geometry of the Grassmannian* $\mathrm{Gr}(p,n)$: The Grassmann manifold (or the Grassmannian) is defined as the set of all $p$-dimensional linear subspaces in $\mathbf{R}^n$ and is denoted by $\mathrm{Gr}(p,n)$, where $p \in \mathbf{Z}^+$, $n \in \mathbf{Z}^+$, $n \geq p$. Grassmannian is a symmetric space and can be identified with the quotient space $\mathrm{SO}(n)/S(O(p) \times O(n-p))$, where $S(O(p) \times O(n-p))$ is the set of all $n \times n$ matrices whose top left $p \times p$ and bottom right $n-p \times n-p$ submatrices are orthogonal and all other entries are 0, and overall the determinant is 1. A point $\mathcal{X} \in \mathrm{Gr}(p,n)$ can be specified by a basis, $X$. We say that $\mathcal{X} = \mathrm{Col}(X)$ if $X$ is a basis of $\mathcal{X}$, where $\mathrm{Col}(\cdot)$ is the column span operator. It is easy to see that the general linear group $\mathrm{GL}(p)$ acts isometrically, freely and properly on $\mathrm{St}(p,n)$. Moreover, $\mathrm{Gr}(p,n)$ can be identified with the quotient space $\mathrm{St}(p,n)/\mathrm{GL}(p)$. Hence, the projection map $\Pi : \mathrm{St}(p,n) \to \mathrm{Gr}(p,n)$ is a *Riemannian submersion*, where $\Pi(X) \triangleq \mathrm{Col}(X)$. Moreover, the triplet $(\mathrm{St}(p,n), \Pi, \mathrm{Gr}(p,n))$ is a fiber bundle.

At every point $X \in \mathrm{St}(p,n)$, we can define the *vertical space*, $\mathcal{V}_X \subset T_X \mathrm{St}(p,n)$ to be $\mathrm{Ker}(\Pi_{*X})$. Further, given $g^{\mathrm{St}}$, we define the *horizontal space*, $\mathcal{H}_X$ to be the $g^{\mathrm{St}}$-orthogonal complement of $\mathcal{V}_X$. Now, from the theory of principal bundles, for every vector field $\widetilde{U}$ on $\mathrm{Gr}(p,n)$, we define the *horizontal lift* of $\widetilde{U}$ to be the unique vector field $U$ on $\mathrm{St}(p,n)$ for which $U_X \in \mathcal{H}_X$ and $\Pi_{*X} U_X = \widetilde{U}_{\Pi(X)}$, for all $X \in \mathrm{St}(p,n)$. As, $\Pi$ is a Riemannian submersion, the isomorphism $\Pi_{*X}|_{\mathcal{H}_X} : \mathcal{H}_X \to T_{\Pi(X)} \mathrm{Gr}(p,n)$ is an isometry from $(\mathcal{H}_X, g^{\mathrm{St}}_X)$ to $(T_{\Pi(X)} \mathrm{Gr}(p,n), g^{\mathrm{Gr}}_{\Pi(X)})$. So, $g^{\mathrm{Gr}}_{\Pi(X)}$ is defined as

$$g^{\mathrm{Gr}}_{\Pi(X)}(\widetilde{U}_{\Pi(X)}, \widetilde{V}_{\Pi(X)}) = g^{\mathrm{St}}_X(U_X, V_X) = \mathrm{trace}\big((X^T X)^{-1} U_X^T V_X\big), \tag{2.4}$$

where, $\widetilde{U}, \widetilde{V} \in T_{\Pi(X)} \mathrm{Gr}(p,n)$ and $\Pi_{*X} U_X = \widetilde{U}_{\Pi(X)}$, $\Pi_{*X} V_X = \widetilde{V}_{\Pi(X)}$, $U_X \in \mathcal{H}_X$ and $V_X \in \mathcal{H}_X$.

**3. Gaussian distribution on homogeneous spaces.** In this section, we define the Gaussian distribution, $\mathcal{N}(\bar{x}, \sigma)$ on a compact homogeneous space, $\mathcal{M}$, $\bar{x} \in \mathcal{M}$ (location parameter), $\sigma > 0$ (scale parameter), and then propose a sampling algorithm to draw samples from the Gaussian distribution on $\mathrm{St}(p,n)$. Furthermore, we will show that the maximum likelihood estimator (MLE) of $\bar{x}$ is the Fréchet mean (FM) [Fréchet (1948)] of the samples.

We define the probability density function, $f(\cdot; \bar{x}, \sigma)$ with respect to $\omega^{\mathcal{M}}$ (the volume form) of the Gaussian distribution $\mathcal{N}(\bar{x}, \sigma)$ on $\mathcal{M}$ as

$$f(x; \bar{x}, \sigma) = \frac{1}{C(\sigma)} \exp\left(\frac{-d^2(x, \bar{x})}{2\sigma^2}\right). \tag{3.1}$$

The above distribution is a valid probability density function, provided that the normalization factor, $C(\sigma)$ is finite and furthermore, is a constant, that is, does not depend on $\bar{x}$ which we will prove next.

PROPOSITION 3.1. *Let us define* $Z(\bar{x}, \sigma) \triangleq \int_{\mathcal{M}} \widetilde{f}(x; \bar{x}, \sigma) \omega^{\mathcal{M}}(x)$, *where* $\widetilde{f}$ *is the kernel of* $f$. $Z(\bar{x}, \sigma)$ *is finite for all compact manifolds.*

PROOF. Observe that the kernel $\widetilde{f}(x; \bar{x}, \sigma) \leq 1$ for all $x$. Hence, $Z(\bar{x}, \sigma) \leq \int_{\mathcal{M}} \omega^{\mathcal{M}}(x) \leq \infty$ as the volume of any compact manifold is finite. □

PROPOSITION 3.2. *With $Z(\bar{x}, \sigma)$ as defined as above, $C(\sigma) = Z(\bar{x}, \sigma) = Z(o, \sigma)$, where $o \in \mathcal{M}$ is the origin.*

PROOF. As the group action on $\mathcal{M}$ is transitive, there exists $g \in G$ s.t., $\bar{x} = g.o.$,

$$\begin{aligned} Z(\bar{x}, \sigma) &= \int_{\mathcal{M}} \widetilde{f}(x; \bar{x}, \sigma) \omega^{\mathcal{M}}(x) \\ &= \int_{\mathcal{M}} \widetilde{f}(g^{-1}.x; g^{-1}.\bar{x}, \sigma) \omega^{\mathcal{M}}(x) \qquad \text{(using Fact 2 in Section 2)} \\ &= \int_{\mathcal{M}} \widetilde{f}(x; o, \sigma) \omega^{\mathcal{M}}(x) \qquad \text{(using Fact 2 in Section 2)} \\ &= Z(o, \sigma). \end{aligned}$$

Hence, $C(\sigma) = Z(o, \sigma)$, that is, does not depend on $\bar{x}$. □

Now that we have a valid definition of a Gaussian distribution, $\mathcal{N}(\bar{x}, \sigma)$ on a compact Homogeneous space, we propose a sampling algorithm for drawing samples from $\mathcal{N}(\bar{X}, \sigma)$ on $\mathrm{St}(p, n)$ (which is a homogeneous space), $\bar{X} \in \mathrm{St}(p, n), \sigma > 0$.

3.1. *Sampling algorithm.* In order to draw samples from $\mathcal{N}(\bar{X}, \sigma)$ on $\mathrm{St}(p, n)$, it is sufficient to draw samples from $\mathcal{N}(O, \sigma)$ where $O \in \mathrm{St}(p, n)$ is the origin. Then, using group operation, we can draw samples from $\mathcal{N}(\bar{X}, \sigma)$ for any $\bar{X} \in \mathrm{St}(p, n)$. We will assume, $O = [I_p\ 0]$ ($n \times p$ matrix with the upper right $p \times p$ block being the identity and the rest being zeros). We will first draw samples from $\mathcal{N}(\mathcal{O}, \sigma)$ on $\mathrm{Gr}(p, n)$, where $\mathcal{O} = \Pi(O)$ and use this sample to get a sample on $\mathrm{St}(p, n)$ using $\mathcal{N}(O, \sigma)$. Note that $\mathrm{Gr}(p, n)$ is a symmetric space, and hence a homogeneous space, and thus we have a valid Gaussian density on $\mathrm{Gr}(p, n)$ using equation (3.1).

PROPOSITION 3.3. *Let $\mathcal{X} \sim \mathcal{N}(\mathcal{O}, \sigma)$ where $\mathcal{O} = \Pi(O)$, $X^T X = I$. Then $\mathrm{Exp}_O(W) \sim \mathcal{N}(O, \sigma)$, with $W = U \Theta V^T$, where $U \Sigma V^T = X(O^T X)^{-1} - O$ and $\Theta = \arctan \Sigma$.*

PROOF. It is sufficient to show that $d(O, \mathrm{Exp}_O(W)) = d(\mathcal{O}, \mathcal{X})$. Recall that $(\mathrm{St}(p, n), \Pi, \mathrm{Gr}(p, n))$ forms a fiber bundle. Moreover, the isomorphism, $\Pi_{*X}|_{\mathcal{H}_X} : \mathcal{H}_X \to T_{\Pi(X)} \mathrm{Gr}(p, n)$, is an isometry from $(\mathcal{H}_X, g_X^{\mathrm{St}})$ to

$(T_{\Pi(X)}\,\mathrm{Gr}(p,n), g^{\mathrm{Gr}}_{\Pi(X)})$, for all $X \in \mathrm{St}(p,n)$. From Absil, Mahony and Sepulchre (2004), we know that $\Pi_{*O}(W) = \mathrm{Exp}^{-1}_{\mathcal{O}}(\mathcal{X})$. So,

$$\begin{aligned} d^2(\mathcal{O},\mathcal{X}) &= g^{\mathrm{Gr}}_{\mathcal{O}}\big(\mathrm{Exp}^{-1}_{\mathcal{O}}(\mathcal{X}), \mathrm{Exp}^{-1}_{\mathcal{O}}(\mathcal{X})\big) \\ &= g^{\mathrm{St}}_{O}(W,W) \\ &\quad \big[\text{as } \Pi_{*O} \text{ is an isomorphism and using equation (2.4)}\big] \\ &= d^2\big(O, \mathrm{Exp}_O(W)\big) \qquad \big[\text{using equation (2.3)}\big]. \end{aligned}$$

□

Using Proposition 3.3, we can generate a sample from $\mathcal{N}(O,\sigma)$ on $\mathrm{St}(p,n)$, using a sample from $\mathcal{N}(\mathcal{O},\sigma)$ on $\mathrm{Gr}(p,n)$. We will now propose an algorithm to draw samples from $\mathcal{N}(\mathcal{O},\sigma)$ on $\mathrm{Gr}(p,n)$. Recall that $\mathrm{Gr}(p,n)$ can be identified as $\mathrm{SO}(n)/S(O(p)\times O(n-p))$ which is a semisimple symmetric space of compact type (let it be denoted by $\mathfrak{g} = \mathfrak{h} + \mathfrak{p}$). Also, recall from Section 2 that every compact semisimple symmetric space has a dual semisimple symmetric space of noncompact type (denoted by $\widetilde{\mathfrak{g}} = \mathfrak{h} + J(\mathfrak{p})$. Here, $\mathfrak{g} = \mathrm{so}(n)$, $\mathfrak{h} = \big[\begin{smallmatrix} \bar{U} & 0 \\ 0 & \bar{V}\end{smallmatrix}\big]$, where $\bar{U} \in \mathrm{so}(p)$, $\bar{V} \in \mathrm{so}(n-p)$, $\mathfrak{p} = \big[\begin{smallmatrix} 0 & \bar{W} \\ -\bar{W}^T & 0\end{smallmatrix}\big]$, $\bar{W} \in \mathbf{R}^{p\times(n-p)}$. Then $\widetilde{\mathfrak{g}} = \mathrm{so}(p,n-p)$, and the corresponding Lie group, denoted by $\widetilde{G} = \mathrm{SO}(p,n-p)$ [with a slight abuse of notation we use $\mathrm{SO}(p,n-p)$ to denote the identity component]. Here, $\mathrm{SO}(p,n-p)$ is the special pseudo-orthogonal group, that is,

$$\begin{aligned} \mathrm{SO}(p,n-p) &\triangleq \big\{\widetilde{g} \mid \widetilde{g} I_{(p,n-p)} \widetilde{g}^T = I_{(p,n-p)}, \det(\widetilde{g}) = 1\big\}, \\ I_{(p,n-p)} &\triangleq \mathrm{diag}(\underbrace{1,\dots,1}_{p \text{ times}}, \underbrace{-1,\dots -1}_{(n-p)\text{ times}}). \end{aligned}$$

Thus, the dual noncompact type symmetric space of $\mathrm{SO}(n)/S(O(p)\times O(n-p))$ [identified with $\mathrm{Gr}(p,n)$] is $\mathrm{SO}(p,n-p)/S(O(p)\times O(n-p))$. Recently, in Said et al. (2016), an algorithm to draw samples from a Gaussian distribution on symmetric spaces of noncompact type was presented. We will use the following proposition to get a sample from a Gaussian distribution on the dual compact symmetric space.

PROPOSITION 3.4. *Let* $\mathcal{X}' \sim \mathcal{N}(\mathcal{O}',\sigma)$. *Let* $\mathcal{X}' = \exp(\mathrm{Ad}(\bar{U})\bar{V}).\mathcal{O}'$, *where* Ad *is the adjoint representation,* $\bar{U} \in \mathfrak{h}$, $\bar{V} \in J(\mathfrak{p})$. *Then* $\mathcal{X} \sim \mathcal{N}(\mathcal{O},\sigma)$, *where* $\mathcal{X} = \exp(\bar{U})\cdot\exp(\widetilde{V}).\mathcal{O}$ *and* $\bar{V} = J(\widetilde{V})$.

PROOF. Observe that $\mathcal{O}' = H = \mathcal{O}$. So, it suffices to show that $d(\mathcal{X}',\mathcal{O}) = d(\mathcal{X},\mathcal{O})$.

$$\begin{aligned} d^2(\mathcal{X}',\mathcal{O}) &= B(\bar{V},\bar{V}) \qquad (\text{the metric corresponds to Killing form } B \text{ on } \mathfrak{p}) \\ &= B\big(J(\widetilde{V}), J(\widetilde{V})\big) \end{aligned}$$

$$= B(\widetilde{V}, \widetilde{V}) \qquad \text{(Killing form is invariant under automorphisms)}$$
$$= d^2(\mathcal{X}, \mathcal{O}). \qquad \square$$

Note that, the mapping $H \times \mathfrak{p} \to G$ given by $(h, \exp(\widetilde{V})) \mapsto h \cdot \exp(\widetilde{V})$ is a diffeomorphism and is used to construct $\mathcal{X}$ from $(\bar{U}, \exp(\widetilde{V}))$. The mapping $\mathcal{X}' = \exp(\mathrm{Ad}(\bar{U})\bar{V}).\mathcal{O}'$ is called the polar coordinate transform. Now, using Propositions 3.3 and 3.4, starting with a sample drawn from a Gaussian distribution on $\mathrm{SO}(p, n-p)/S(O(p) \times O(n-p))$, we get a sample from Gaussian distribution on $\mathrm{St}(p, n)$. We would like to point out that we do not have to compute the normalizing constant explicitly in order to draw samples, because in order to get samples on $\mathrm{SO}(p, n-p)/S(O(p) \times O(n-p))$, we can draw samples using the Algorithm 1 in Said et al. (2016), which draws samples from the kernel of the density.

3.2. *Maximum likelihood estimation* (*MLE*) *of* $\bar{X}$. Let, $X_1, X_2, \dots X_N$ be i.i.d. samples drawn from $\mathcal{N}(\bar{X}, \sigma)$ with bounded support (described subsequently) on $\mathrm{St}(p, n)$, for some $\bar{X} \in \mathrm{St}(p, n)$, $\sigma > 0$. Then, by Proposition 3.6, the MLE of $\bar{X}$ is the Fréchet mean (FM) [Fréchet (1948)] of $\{X_i\}_{i=1}^N$. Fréchet mean (FM) [Fréchet (1948)] of $\{X_i\}_{i=1}^N \subset \mathrm{St}(p, n)$ is defined as follows:

$$M = \operatorname*{arg\,min}_{X \in \mathrm{St}(p,n)} \sum_{i=1}^{N} d^2(X_i, X). \tag{3.2}$$

We define an (open) "geodesic ball" of radius $r > 0$ to be $\mathcal{B}(X, r) = \{X_i | d(X, X_i) < r\}$ s.t., there exists a length minimizing geodesic between $X$ to any $X_i \in \mathcal{B}(X, r)$. A "geodesic ball" is said to be "regular" *iff* $r < \pi/2(\sqrt{\kappa})$, where $\kappa$ is the maximum sectional curvature. The existence and uniqueness of the Fréchet mean (FM) is ensured *iff* the support of the distribution $\mathcal{N}(\bar{X}, \sigma)$ is within a regular geodesic ball [Afsari (2011), Kendall (1990)].

PROPOSITION 3.5. *Let* $X \in \mathrm{St}(p, n)$, $U, V \in \mathcal{H}_X$, *then* $0 \leq \kappa(U, V) \leq 2$.

PROOF. Let, $\mathcal{X} = \Pi(X)$. Then there exists a unique $\widetilde{U}, \widetilde{V} \in T_{\mathcal{X}} \mathrm{Gr}(p, n)$ s.t. $\widetilde{U} = \Pi_{*X} U$, $\widetilde{V} = \Pi_{*X} V$. $0 \leq \kappa(\widetilde{U}, \widetilde{V}) \leq 2$ [Wong (1968)]. Now, using O'Neil's formula [Cheeger and Ebin (1975)], we know that

$$\kappa(\widetilde{U}, \widetilde{V}) = \kappa(U, V) + \frac{3}{4} \| \mathrm{vert}_X([U, V]) \|^2,$$

where $\mathrm{vert}_X$ is the orthogonal projection operator on $\mathcal{V}_X$. Clearly, as the second term in the above summation is nonnegative and $\kappa(U, V)$ is nonnegative [as $\mathrm{St}(p, n)$ is of compact type], the result follows. $\square$

Observe that the support of $\mathcal{N}(\bar{X}, \sigma)$ as defined in Proposition 3.3 is a subset of $\mathcal{H} \triangleq \bigcup_X \mathrm{Exp}_X(\mathcal{H}_X) \subset \mathrm{St}(p, n)$, $\mathcal{H}$ is an arbitrary union of open sets, and hence is open. Thus, we can give $\mathcal{H}$ a manifold structure and using Proposition 3.5, we can say that if the support of $\mathcal{N}(\bar{X}, \sigma)$ is within a geodesic ball $\mathcal{B}(\bar{X}, \pi/2(\sqrt{2}))$, FM exists and is unique. For the rest of the paper, we assume this condition to ensure the existence and uniqueness of FM.

PROPOSITION 3.6. *Let $X_1, X_2, \dots, X_N$ be i.i.d. samples drawn from $\mathcal{N}(\bar{X}, \sigma)$ on* $\mathrm{St}(p, n)$ [*support of $\mathcal{N}(\bar{X}, \sigma)$ is within a geodesic ball $\mathcal{B}(\bar{X}, \pi/2(\sqrt{2}))$*], $\sigma > 0$. *Then the MLE of $\bar{X}$ is the FM of $\{X_i\}$.*

PROOF. The likelihood of $\bar{X}$ given the i.i.d. samples $\{X_i\}$ is given by

$$L(\bar{X}, \sigma; \{X_i\}_{i=1}^N) = \frac{1}{C(\sigma)} \prod_{i=1}^{N} \exp\left(\frac{-d^2(X_i, \bar{X})}{2\sigma^2}\right), \tag{3.3}$$

where $C(\sigma)$ is defined as in equation (3.1). Now, maximizing log-likelihood function with respect to $\bar{X}$ is equivalent to minimizing $\sum_{i=1}^N d^2(X_i, \bar{X})$ with respect to $\bar{X}$. This gives the MLE of $\bar{X}$ to be the FM of $\{X_i\}_{i=1}^N$ as can be verified using equation (3.2). □

**4. Inductive Fréchet mean on the Stiefel manifold.** In this section, we present an inductive formulation for computing the Fréchet mean (FM) [Fréchet (1948), Karcher (1977)] on the Stiefel manifold. We also prove the *weak consistency* of our FM estimator on the Stiefel manifold.

*Algorithm for inductive Fréchet mean estimator*

Let $X_1, X_2, \dots$ be i.i.d. samples drawn from $\mathcal{N}(\bar{X}, \sigma)$ [whose support is within a geodesic ball $\mathcal{B}(\bar{X}, \pi/2(\sqrt{2}))$] on $\mathrm{St}(p, n)$. Then we define the inductive FM estimator (*StiFME*) $M_k$ by the recursion in equations (4.1) and (4.2).

$$M_1 = X_1, \tag{4.1}$$

$$M_{k+1} = \Gamma_{M_k}^{X_{k+1}}(\omega_{k+1}), \tag{4.2}$$

where $\Gamma_X^Y : [0, 1] \to \mathrm{St}(p, n)$ is the geodesic from $X$ to $Y$ defined as $\Gamma_X^Y(t) := \mathrm{Exp}_X(t\,\mathrm{Exp}_X^{-1}(Y))$ and $\omega_{k+1} = \frac{1}{k+1}$. Equation (4.2) simply means that the $(k + 1)$th estimator lies on the geodesic between the $k$th estimate and the $(k + 1)$th sample point. This simple inductive estimator can be shown to converge to the Fréchet expectation, that is, $\bar{X}$, as stated in Theorem 4.1.

THEOREM 4.1. *Let, $X_1, X_2 \dots X_N$ be i.i.d. samples drawn from a Gaussian distribution $\mathcal{N}(\bar{X}, \sigma)$ on* $\mathrm{St}(p, n)$ (*with a support inside a regular geodesic ball of radius $< \pi/2\sqrt{2}$*). *Then the inductive FM estimator* (StiFME) *of these samples, that is, $M_N$ converges to $\bar{X}$ as $N \to \infty$.*

PROOF. We will start by first stating the following propositions.

PROPOSITION 4.1. *Using Proposition 2.5, we know that* $\Pi : \mathrm{SO}(n) \to \mathrm{SO}(n)/\mathrm{SO}(n-p)$ *is a principal bundle, and moreover, using Proposition 2.6, we know that this map is a covering map in the neighborhood of* $\mathrm{SO}(n-p)$ *in* $\mathrm{SO}(n)/\mathrm{SO}(n-p)$. *Let* $g^{\mathrm{SO}}$ *be the Riemannian metric on* $\mathrm{SO}(n)$ *and* $g^q$ *be the metric on the quotient space* $\mathrm{SO}(n)/\mathrm{SO}(n-p)$. *Then* $g^{\mathrm{SO}} = \Pi^* g^q$.

PROPOSITION 4.2. *Let,* $X_i = g_i H$, *where* $H := \mathrm{SO}(n-p)$ *and* $g_i \in G := \mathrm{SO}(n)$. *Let* $M$ *is an defined in equation (3.2), then* $M = g_M H$, *where* $g_M = \arg\min_{g \in \mathrm{SO}(n)} \sum_{i=1}^{N} d^2(g_i, g)$.

PROOF. Let $M = \bar{g}H$, for some $\bar{g} \in G$. Then observe that

$$\begin{aligned} d^2(X_i, M) &= d^2(g_i H, \bar{g}H) \\ &= d^2(\bar{g}^{-1} g_i H, H) \qquad \text{using property 2 of homogeneous space} \\ &= d^2(\bar{g}^{-1} g_i, e) \qquad \text{using Proposition 4.1} \\ &= d^2(g_i, \bar{g}) \qquad \text{as } \mathrm{SO}(n) \text{ a Lie group.} \end{aligned}$$

Thus the claim holds. □

By Proposition 4.2, we can see that in order to prove Theorem 4.1, it is sufficient to show weak consistency on $\mathrm{SO}(n)$. We will state and prove the weak consistency on $\mathrm{SO}(n)$ in the next theorem. □

THEOREM 4.2. *Using the hypothesis in Theorem 4.1, let* $g_1, g_2, \ldots, g_N$ *be the corresponding i.i.d. samples drawn from the (induced) Gaussian distribution* $\mathcal{N}(\bar{g}, \sigma)$ *on* $\mathrm{SO}(n)$ *where* $\bar{X} = \bar{g}H$ $[H := \mathrm{SO}(n-p)]$ *[it is easy to show using Proposition 4.2 that this (induced) distribution on* $\mathrm{SO}(n)$ *is indeed a Gaussian distribution on* $\mathrm{SO}(n)$*]. Then the inductive FM estimator* (StiFME) *of these samples, that is,* $g_N$ *converges to* $\bar{g}$ *as* $N \to \infty$.

PROOF. Since $\mathrm{SO}(n)$ is a special case of the (compact) Stiefel manifold, that is, when $p = n - 1$ [as $\mathrm{SO}(n)$ can be identified with $\mathrm{St}(n-1, n)$], we will use $X$ instead of $g$ for notational simplicity. Let $X \in \mathrm{SO}(n)$. Any point in $\mathrm{SO}(n)$ can be written as a product of $n(n-1)/2$ planar rotation matrices by the following claim.

PROPOSITION 4.3. *Any arbitrary element of* $\mathrm{SO}(n)$ *can be written as the composition of planar rotations in the planes generated by the* $n$ *standard orthogonal basis vectors of* $\mathbf{R}^n$.

PROOF. The proof is straightforward. Moreover, each element of $\mathrm{SO}(n)$ is a product of $n(n-1)/2$ planar rotations. □

By virtue of Proposition 4.3, we can express $X$ as a product of $n(n-1)/2$ planar rotation matrices. Each planar rotation matrix can be mapped onto $\mathbf{S}^{n-1}$, hence ∃ diffeomorphism $F : \mathrm{SO}(n) \to \underbrace{\mathbf{S}^{n-1} \times \cdots \times \mathbf{S}^{n-1}}_{n(n-1)/2 \text{ times}}$. Let us denote this product space of hyperspheres by $\mathfrak{O}(n-1, \frac{n(n-1)}{2})$. Then $F$ is a diffeomorphism from $\mathrm{SO}(p)$ to $\mathfrak{O}(n-1, \frac{n(n-1)}{2})$. Let $g^{\mathfrak{O}}$ be a Riemannian metric on $\mathfrak{O}(n-1, \frac{n(n-1)}{2})$. Let $\nabla^{\mathfrak{O}}$ be the Levi–Civita connection on $T\mathfrak{O}(n-1, \frac{n(n-1)}{2})$. Since $F$ is a diffeomorphism, every vector field $U$ on $\mathrm{SO}(n)$ pushes forward to a well-defined vector field $F_*U$ on $\mathfrak{O}(n-1, \frac{n(n-1)}{2})$. Define a map

$$\nabla^{\mathrm{SO}} : \Xi(\mathrm{TSO}(n)) \times \Xi(\mathrm{TSO}(n)) \to \Xi(\mathrm{TSO}(n)),$$
$$(U, V) \mapsto \nabla^{\mathrm{SO}}_U V,$$

where $\Xi(\mathrm{TSO}(n))$ gives the section of $\mathrm{TSO}(n)$.

PROPOSITION 4.4. *$\nabla^{\mathrm{SO}}$ is the Levi–Civita connection on $\mathrm{SO}(n)$ equipped with the pull-back Riemannian metric $F^*g^{\mathfrak{O}}$.*

PROPOSITION 4.5. *Given the hypothesis and the notation as above, if $\gamma$ is a geodesic on $\mathrm{SO}(n)$, $F \circ \gamma$ is a geodesic on $\mathfrak{O}(n-1, \frac{n(n-1)}{2})$.*

PROOF. Let, $\hat{\gamma} = F \circ \gamma$ be a curve in $\mathcal{O}(n-1, \frac{n(n-1)}{2})$. Then

$$0 = F_*0 = F_*(\nabla^{\mathrm{SO}}_{\gamma'}\gamma') = F_*(F_*^{-1}(\nabla^{\mathfrak{O}}_{F_*\gamma'}F_*\gamma'))$$
$$= \nabla^{\mathfrak{O}}_{\hat{\gamma}'}\hat{\gamma}'.$$

Hence, $\hat{\gamma}$ is a geodesic on $\mathcal{O}(n-1, \frac{n(n-1)}{2})$. □

Now, analogous to equation (4.1), we can define the FM estimator on $\mathrm{SO}(n)$ where the geodesic, $\Gamma_{M_k}^{X_{k+1}}(\omega_{k+1}) = \mathrm{Exp}_{M_k}(\omega_{k+1}\,\mathrm{Exp}_{M_k}^{-1}(X_{k+1}))$. Note that on $\mathrm{SO}(n)$, $\mathrm{Exp}_{M_k}(\omega_{k+1}\,\mathrm{Exp}_{M_k}^{-1}(X_{k+1})) = M_k \exp(\omega_{k+1}\log(M_k^{-1}X_{k+1}))$.

PROPOSITION 4.6. $F_*\,\mathrm{Exp}_{M_k}^{-1}(X_{k+1}) = \mathrm{Exp}_{F(M_k)}^{-1}(F(X_{k+1}))$.

PROOF. Let $\gamma : [0,1] \to \mathrm{SO}(n)$ be a geodesic from $M_k$ to $X_{k+1}$. Then $\mathrm{Exp}_{M_k}^{-1}(X_{k+1}) = \frac{d}{dt}(\gamma(t))|_{t=0}$. Using Proposition 4.5, $F \circ \gamma$ is a geodesic from

$F(M_k)$ to $F(X_{k+1})$:

$$\begin{aligned}\mathrm{Log}_{F(M_k)} F(X_{k+1}) &= \frac{d}{dt}\big(F \circ \gamma(t)\big)\Big|_{t=0} \\ &= F_* \frac{d}{dt}\big(\gamma(t)\big)\Big|_{t=0} \\ &= F_* \mathrm{Exp}_{M_k}^{-1}(X_{k+1}). \qquad \square\end{aligned}$$

Let $\bar{U} = \mathrm{Exp}_{F(M_k)}^{-1}(F(X_{k+1}))$ and $\hat{U} = \mathrm{Exp}_{M_k}^{-1}(X_{k+1})$. Using Proposition 4.6, we get

$$\begin{aligned}g^{\mathrm{SO}}(\hat{U}, \hat{U}) &= F^* g^{\mathfrak{O}}(\hat{U}, \hat{U}) \\ &= g^{\mathfrak{O}}(F_*\hat{U}, F_*\hat{U}) \\ &= g^{\mathfrak{O}}(\bar{U}, \bar{U}).\end{aligned}$$

Thus, in order to show weak consistency of our proposed estimator on $\{g_i\} \subset \mathrm{SO}(n)$, it is sufficient to show the weak consistency of our estimator on $\{F(g_i)\} \subset \mathfrak{O}(n-1, \frac{n(n-1)}{2})$. A proof of the weak consistency of our proposed FM estimator on the hypersphere has been shown in Salehian et al. (2015) (which can be trivially extended to the product of hyperspheres). This proof of weak consistency on the hypersphere in turn proves the weak consistency on $\mathrm{SO}(n)$. $\square$

Since we have now shown that our proposed FM estimator on $\mathrm{St}(p,n)$ is (weakly) consistent, we claim that $\mathrm{Var}(M_N) \geq \mathrm{Var}(\widehat{M}_N)$ as $N \to \infty$, where $\widehat{M}_N$ is the MLE of $\bar{X}$ when $\{X_i\}_{i=1}^N$ are i.i.d. samples from $\mathcal{N}(\bar{X}, \sigma)$ on $\mathrm{St}(p,n)$. The following proposition computes the Fisher information of $\bar{X}$ when samples are drawn from $\mathcal{N}(\bar{X}, \sigma)$ on $\mathrm{St}(p,n)$.

PROPOSITION 4.7. *Let* $\mathbf{X}$ *be a random variable which follows* $\mathcal{N}(\bar{X}, \sigma)$ *on* $\mathrm{St}(p,n)$. *Then* $I(\bar{X}) = 1/\sigma^2$.

PROOF. The likelihood of $\bar{X}$ is given by

$$L(\bar{X}; \sigma, \mathbf{X} = X) = \frac{1}{C(\sigma)} \exp\left(\frac{-d^2(X, \bar{X})}{2\sigma^2}\right). \tag{4.3}$$

Then $I(\bar{X}) = E_{\mathbf{X}}[\langle \frac{\partial l}{\partial \bar{X}}, \frac{\partial l}{\partial \bar{X}} \rangle_{\bar{X}}]$, where $l(\bar{X}; \sigma, X)$ is the log likelihood. Now, $l(\bar{X}; \sigma, X) = \frac{\mathrm{Exp}_{\bar{X}}^{-1} X}{\sigma^2}$, hence $E_{\mathbf{X}}[\langle \frac{\partial l}{\partial \bar{X}}, \frac{\partial l}{\partial \bar{X}} \rangle] = E_{\mathbf{X}}[\langle \mathrm{Exp}_{\bar{X}}^{-1} X, \mathrm{Exp}_{\bar{X}}^{-1} X \rangle_{\bar{X}}] = E_{\mathbf{X}}[d^2(X, \bar{X})]$. Now, observe that $\mathrm{Var}(\mathbf{X}) = E_{\mathbf{X}}[d^2(X, \bar{X})]$ [here, definition of variance of a manifold valued random variable is as in Pennec (2006)], where from the definition of the Gaussian distribution, $\mathrm{Var}(\mathbf{X}) = \sigma^2$. Hence, $I(\bar{X}) = 1/\sigma^2$. $\square$

As $\mathrm{Var}(\widehat{M}_N) = \sigma^2$ (as we have shown that $\widehat{M}_N$ is the FM of the samples in Proposition 3.6) when the number of samples tends to infinity, and $\sigma^2 = 1/I(\bar{X})$ by Proposition 4.7, we conclude that MLE achieves the Cramér–Rao lower bound asymptotically (this observation is in line with the normal random vector). As we have shown, consistency of our estimator, hence $\mathrm{Var}(M_N)$ is lower bounded by $\mathrm{Var}(\widehat{M}_N)$ as $N \to \infty$. In other words, asymptotically $\mathrm{Var}(M_N) \geq \mathrm{Var}(\widehat{M}_N) = \sigma^2$.

**5. Experimental results.** In this section, we present experiments demonstrating the performance of *StiFME* in comparison to the batch mode counterpart with "warm start"(which uses the gradient descent on the sum of squared geodesic distances cost function, henceforth termed *StFME*) on synthetic and real datasets. By "warm start" we mean that when a new data point is acquired as input, we initialize the FM to its computed value prior to the arrival/acquisition of the new data point. All the experimental results reported here were performed on a desktop with a 3.33 GHz Intel-i7 CPU with 24 GB RAM.

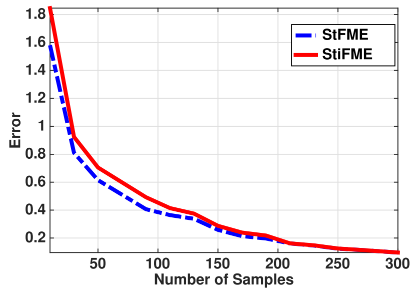
5.1. *Comparative performance of* StiFME *on synthetic data.* We generated 1000 i.i.d. samples drawn from a Normal distribution on $\mathrm{St}(p, n)$ with variance 0.25 and expectation $\tilde{I}$, where

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

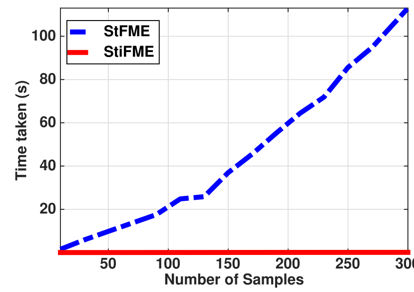
We input these i.i.d. samples to both *StiFME* and *StFME*. To compare the performance, we compute the error, which is the distance [on $\mathrm{St}(p, n)$] between the computed FM and the known true FM $\tilde{I}$. We also report the computation time for both these cases. We performed this experiment 5000 times and report the average error and the average computation time. The comparison plot for the average error is shown in Figure 1(a); here, $n = 50$, $p = 10$. In order to achieve faster convergence of *StFME*, we used the "warm start" technique, that is, FM of $k$ samples is used to initialize the FM computation for $k + 1$ samples. From this plot, it is evident that the average accuracy error of *StiFME* is almost same as that of *StFME*.

The computation time comparison between *StiFME* and *StFME* is shown in Figure 1(a). From this figure, we can see that *StiFME* outperforms *StFME*. As the number of samples increases, the computational efficiency of *StiFME* over *StFME* becomes significantly large. We can also see that the time requirement for *StiFME* is almost constant with respect to the the number of samples, which makes *StiFME* computationally very efficient and attractive for a large number of data samples.

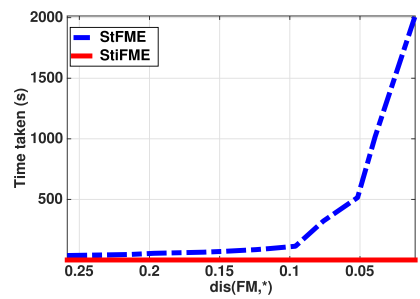
Another interesting question to ask is: how much computation time is needed in order to estimate the FM with a given error tolerance? We answer this question through the plot in Figure 1(c) and present a comparison of the time required for *StiFME* and *StFME*, respectively, to reach the specified error tolerance. From Figure 1(c), it is evident that the time required to reach the specified error tolerance by *StiFME* is far less than that required by *StFME*.




(a) Average error.




(b) Average running time.




(c) Time required to attain a specified accuracy.

FIG. 1. *Comparison between* StFME *and* StiFME.

5.2. *Clustering action data from videos.* In this subsection, we applied our FM estimator to cluster the KTH video action data [Schuldt, Laptev and Caputo (2004)]. This data contains 6 actions performed by 25 human subjects in 4 scenarios (denoted by "d1," "d2," "d3" and "d4"). From each video, we extracted a sequence of frames. Then, from each frame we computed the *Histogram of Oriented Gradients* (HOG) [Dalal and Triggs (2005)] features. We then used an autoregressive moving average (ARMA) model [Doretto (2003)] to model each activity. The equations for the ARMA model are given below:

$$f(t) = Cz(t) + w(t),$$

$$z(t+1) = Az(t) + v(t),$$

where $w$ and $v$ are zero-mean Gaussian noise, $f$ is the feature vector, $z$ is the hidden state, $A$ is the transition matrix and $C$ is the measurement matrix. In Doretto (2003), authors proposed a closed-form solution for $A$ and $C$ by stacking feature vectors over time and performing a singular value decomposition on the feature matrix. More specifically, let $T$ be the number of frames and let $F$ be the matrix formed by stacking the feature vectors from each frame. Let $U\Sigma V^T$ be SVD of $F$, then $A$ and $C$ can be approximated as $C = U$, $A = \Sigma V^T D_1 V(V^T D_2 V)^{-1}\Sigma^{-1}$,

TABLE 1
*Comparison results on the KTH action recognition database*

| **Method** | **Scenario** | **Precision (%)** | **Time (s)** |
|---|---|---|---|
| StFME | d1 | **78.21** | 204.59 |
| StiFME | d1 | 77.33 | **2.32** |
| StFME | d2 | **73.33** | 253.15 |
| StiFME | d2 | 70.67 | **2.48** |
| StFME | d3 | **79.67** | 267.40 |
| StiFME | d3 | 77.91 | **2.59** |
| StFME | d4 | **83.83** | 216.27 |
| StiFME | d4 | 90.73 | **2.82** |

where $D_1$ and $D_2$ are zero matrices with identity in the bottom-left and top-left submatrix, respectively. Clearly, $C$ lies on a Stiefel manifold, but in general $A$ does not have any special structure. Hence, we identify each activity with a product space of $U$, $\Sigma$ and $V$. Note that both $U$ and $V$ lie on the Steifel manifold (possibly of different dimensions) and $\Sigma$ lies in the Euclidean space.

Here, we perform clustering of the actions by doing clustering on the product manifold of $\mathrm{St}(p,n) \times \mathrm{St}(n,n) \times \mathbf{R}^n$. The accuracy is reported in Table 1. From this table, we can see *StiFME* depicts a significant gain in computation time over *StFME* and is comparable in accuracy.

We would like to point out that in the real data experiment, one can easily fit a half-normal distribution on $\{d(X_i, \bar{X})\}$ by viewing the relation of our definition of Gaussian distribution with the kernel of the half-normal distribution on $\{d(X_i, \bar{X})\}$ with location parameter 0 and scale parameter $\sigma^2$. So, the goodness-of-fit can be evaluated using the Chi-squared test where the null hypothesis $H_0$ is that $\{d(X_i, \bar{X})\}$ are drawn from a half-normal distribution.

In this experiment, we estimated the goodness-of-fit in fitting a Gaussian to the set of samples, $\{U_i\}$ (samples collected from a given action), using the aforementioned procedure. We found that the Chi-squared test does not reject the null hypothesis with a 5% significance level, implying that $\{U_i\}$ are indeed drawn from a Gaussian distribution on $\mathrm{St}(p,n)$. We also tried to fit a Gaussian to the entire data, that is, over all actions, and found that the entire data are not drawn from a Gaussian distribution. This is not surprising, as the entire dataset probably follow a mixture of Gaussians as each individual action/ cluster follows a Gaussian distribution.

5.3. *Experiments on vector-cardiogram dataset*. This data set [Downs, Liebman and Mackay (1971)] summarizes vector-cardiograms of 98 healthy children aged between 2–19. Each child has two vector-cardiograms, using the Frank

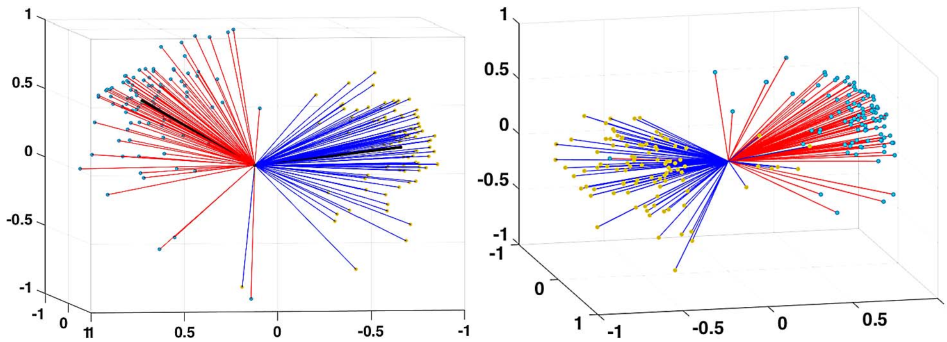



FIG. 2. *Averaging on Vector-cardiogram data*. Data with FM shown in black (*Left*), reconstructed data (*Right*).

and McFee system, respectively. The two vector-cardiograms are represented as two mutually orthogonal orientations in $\mathbf{R}^3$; hence, each vector-cardiogram can be mapped to a point on $\mathrm{St}(2,3)$. We perform statistical analysis via principal geodesic analysis (PGA) [Fletcher et al. (2004)] of the data depicted in Figure 2 (at the top). One of the key steps in PGA is to find the FM, which is depicted in the plot (in black). Further, we reconstructed the data from the first two principal directions (which accounts for $>90\%$ of the data variance) and the reconstructed results are shown in the rightmost plot. The reconstruction error is on the average 0.05 per subject, which implies that the reconstruction is quite accurate.

5.4. *Comparison with stochastic gradient descent based FM estimator.* In this subsection, we present a comparison between *StiFME* and the stochastic gradient descent based FM estimator in Bonnabel (2013).

There are two key differences between the algorithm in Bonnabel (2013) and *StiFME*. As in any stochastic gradient scheme, the next point, that is, $z_t$ in $w_{t+1} = \exp_{w_t}(-\gamma_t H(z_t, w_t))$ [equation (2) in Bonnabel (2013)] is chosen randomly from the given sample set. Hence, the stochastic formulation needs several passes over the sample set and reports the expected value over the passes as the estimated FM. In contrast, *StiFME* is a deterministic algorithm, and hence does not need multiple passes over the data. Moreover, our selection of this weight is primarily in spirit the same as the weights in a recursive arithmetic mean computation in Euclidean space. In contrast, Bonnabel (2013) does not specify any scheme to choose the proper step size $\gamma_t$ [equation (2) of Bonnabel (2013)]. Note that, like in any gradient descent, the algorithm in Bonnabel (2013) is very much dependent on a proper step size selection. Step size selection in gradient descent and its relatives is a hard problem and the most widely used method (Armijo rule) is computationally expensive. We now provide two experimental comparisons with the algorithm in Bonnabel (2013). Consider a data set of 100 samples drawn from a log-Normal distribution, with a small variance of 0.05 on $\mathrm{St}(10,50)$. The distance between

(FM and *StiFME*] and [FM and computed FM using the algorithm in Bonnabel (2013)] (assessed in one pass over the data) are 0.00025 and 0.009, respectively. However, Bonnabel (2013) requires 19 passes over the data to achieve the tolerance of 0.00025 obtained by *StiFME*. For a larger data variance of 0.29 on $\mathrm{St}(10, 50)$, the distance between FM, *StiFME* and FM computed from Bonnabel (2013) are 0.00039 and 0.03 (in one pass over the data), respectively, which is a significant difference. Furthermore, the method in Bonnabel (2013) needs 58 passes over the data to achieve the tolerance achieved by *StiFME*. This clearly indicates better computational efficiency of *StiFME* over the FM estimator in Bonnabel (2013).

5.5. *Time complexity comparison.* The complexity of *StFME* is $\mathcal{O}(\iota N)$, $N$ is the number of samples in the data and $\iota$ is the number of iterations required for convergence. The number of iterations however depends on the step size used; too small a step size causes very slow convergence and too large a step size overshoots the FM. In contrast, the complexity of *StiFME* is $\mathcal{O}(N)$ because it outputs the estimated FM in a single pass through the data. On the other hand, the SGD algorithm proposed in Bonnabel (2013) takes $\mathcal{O}(b\hat{\iota})$, where $b$ is the batch size and $\hat{\iota}$ is the number of iterations to convergence. So, in comparison, *StiFME* is much faster than the other two competing algorithms.

**6. Conclusions.** In this paper, we defined a Gaussian distribution on a compact Riemannian homogenous space and proved that the MLE of the location parameter of this Gaussian distribution yields the FM of the samples drawn from the distribution. Further, we presented a sampling algorithm to draw samples from this Gaussian distribution on the Stiefel manifold (which is a homogeneous space) and a novel recursive estimator, *StiFME*, for computing the FM of these samples. A proof of weak consistency of *StiFME* was also presented. Further, we also showed that the MLE of the location parameter of the Gaussian distribution on $\mathrm{St}(p, n)$ asymptotically achieves the Cramér–Rao lower bound, and hence is efficient. The salient feature of *StiFME* is that it does not require any optimization unlike the traditional methods that seek to optimize the Fréchet functional via gradient descent. This leads to significant savings in computation time and makes it attractive for online applications of FM computation for manifold-valued data, such as clustering, etc. We presented several experiments demonstrating the superior performance of *StiFME* over gradient-descent based competing FM-estimators on synthetic and real data sets.

## REFERENCES

ABSIL, P.-A., MAHONY, R. and SEPULCHRE, R. (2004). Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Appl. Math.* **80** 199–220. MR2035507

AFSARI, B. (2011). Riemannian $L^p$ center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346

ANDO, T., LI, C.-K. and MATHIAS, R. (2004). Geometric means. *Linear Algebra Appl*. **385** 305–334. MR2063358

ARNAUDON, M., BARBARESCO, F. and YANG, L. (2013). Riemannian medians and means with applications to radar signal processing. *IEEE J. Sel. Top. Signal Process*. **7** 595–604.

BHATIA, R. (1997). *Matrix Analysis*. *Graduate Texts in Mathematics* **169**. Springer, New York. MR1477662

BHATTACHARYA, A. and BHATTACHARYA, R. (2008). Statistics on Riemannian manifolds: Asymptotic distribution and curvature. *Proc. Amer. Math. Soc*. **136** 2959–2967. MR2399064

BONNABEL, S. (2013). Stochastic gradient descent on Riemannian manifolds. *IEEE Trans. Automat. Control* **58** 2217–2229. MR3101606

CETINGUL, H. E. and VIDAL, R. (2009). Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds. In *Computer Vision and Pattern Recognition*, 2009. *CVPR* 2009. *IEEE Conference on* 1896–1902. IEEE, NEw York.

CHAKRABORTY, R., BANERJEE, M. and VEMURI, B. (2017). Statistics on the space of trajectories for longitudinal data analysis. In *IEEE International Symposium on Biomedical Imaging*.

CHAKRABORTY, R. and VEMURI, B. C. (2015). Recursive Frechet mean computation on the Grassmannian and its applications to computer vision. In *The IEEE International Conference on Computer Vision* (*ICCV*).

CHARFI, M., CHEBBI, Z., MOAKHER, M. and VEMURI, B. C. (2013). Bhattacharyya median of symmetric positive-definite matrices and application to the denoising of diffusion-tensor fields. In *Biomedical Imaging* (*ISBI*), 2013 *IEEE* 10*th International Symposium on* 1227–1230. IEEE, New York.

CHEEGER, J. and EBIN, D. G. (1975). *Comparison Theorems in Riemannian Geometry*. *North-Holland Mathematical Library* **9**. North-Holland, Amsterdam. MR0458335

CHENG, G. and VEMURI, B. C. (2013). A novel dynamic system in the space of SPD matrices with applications to appearance tracking. *SIAM J. Imaging Sci*. **6** 592–615. MR3035470

CHIKUSE, Y. (1991). Asymptotic expansions for distributions of the large sample matrix resultant and related statistics on the Stiefel manifold. *J. Multivariate Anal*. **39** 270–283. MR1147122

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. *Princeton Mathematical Series* **9**. Princeton Univ. Press, Princeton, NJ. MR0016588

DALAL, N. and TRIGGS, B. (2005). Histograms of oriented gradients for human detection. In *CVPR* **1** 886–893.

DORETTO, G., CHIUSO, A., WU, Y. N. and SOATTO, S. (2003). Dynamic textures. *Int. J. Comput. Vis*. **51** 91–109.

DOWNS, T., LIEBMAN, J. and MACKAY, W. (1971). Statistical methods for vectorcardiogram orientations. In *Vectorcardiography* 2: *Proc. XIth International Symp. on Vectorcardiography*.

FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vis*. **105** 171–185. MR3104017

FLETCHER, P. T. and JOSHI, S. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Process*. **87** 250–262.

FLETCHER, P. T., LU, C., PIZER, S. M. and JOSHI, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imag*. **23** 995–1005.

FRAIKIN, C., HÜPER, K. and DOOREN, P. V. (2007). Optimization over the Stiefel manifold. *PAMM* **7** 1062205–1062206.

FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré* **10** 215–310. MR0027464

GOODALL, C. R. and MARDIA, K. V. (1999). Projective shape analysis. *J. Comput. Graph. Statist*. **8** 143–168. MR1706381

GRENANDER, U. (2008). *Probabilities on Algebraic Structures*. Dover Publications, Mineola, NY.

GROISSER, D. (2004). Newton's method, zeroes of vector fields, and the Riemannian center of mass. *Adv. in Appl. Math*. **33** 95–135. MR2064359

HARTLEY, R., TRUMPF, J., DAI, Y. and LI, H. (2013). Rotation averaging. *Int. J. Comput. Vis.* **103** 267–305. MR3067537

HAUBERG, S., FERAGEN, A. and BLACK, M. J. (2014). Grassmann averages for scalable robust PCA. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3810–3817.

HELGASON, S. (1978). *Differential Geometry, Lie Groups, and Symmetric Spaces. Pure and Applied Mathematics* **80**. Academic Press, New York. MR0514561

HENDRIKS, H. and LANDSMAN, Z. (1998). Mean location and sample mean location on manifolds: Asymptotics, tests, confidence regions. *J. Multivariate Anal.* **67** 227–243. MR1659156

HO, J., CHENG, G., SALEHIAN, H. and VEMURI, B. (2013). Recursive Karcher expectation estimators and geometric law of large numbers. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics* 325–332.

KANEKO, T., FIORI, S. and TANAKA, T. (2013). Empirical arithmetic averaging over the compact Stiefel manifold. *IEEE Trans. Signal Process.* **61** 883–894. MR3018336

KARCHER, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30** 509–541. MR0442975

KENDALL, W. S. (1990). Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence. *Proc. Lond. Math. Soc.* (3) **61** 371–406. MR1063050

LUI, Y. M. (2012). Advances in matrix manifolds for computer vision. *Image Vis. Comput.* **30** 380–388.

MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics*. Wiley, Chichester. Revised reprint of *Statistics of directional data* by Mardia [ MR0336854 (49 #1627)]. MR1828667

MOAKHER, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **26** 735–747. MR2137480

MOAKHER, M. (2006). On the averaging of symmetric positive-definite tensors. *J. Elasticity* **82** 273–296. MR2231065

PATRANGENARU, V. and MARDIA, K. V. (2003). Affine shape analysis and image analysis. In *22nd Leeds Annual Statistics Research Workshop*.

PENNEC, X. (2006). Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vision* **25** 127–154. MR2254442

PENNEC, X., FILLARD, P. and AYACHE, N. (2006). A Riemannian framework for tensor computing. *Int. J. Comput. Vis.* **66** 41–66.

PHAM, D.-S. and VENKATESH, S. (2008). Robust learning of discriminative projection for multicategory classification on the Stiefel manifold. In *Computer Vision and Pattern Recognition*, 2008. *CVPR* 2008. *IEEE Conference on* 1–7. IEEE, New York.

RAO, C. R. (1987). Differential metrics in probability spaces. *Differential Geometry in Statistical Inference* **10** 217–240.

RAO, C. R. (1992). Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in Statistics* 235–247. Springer, New York.

SAID, S., HAJRI, H., BOMBRUN, L. and VEMURI, B. C. (2016). Gaussian distributions on Riemannian symmetric spaces: Statistical learning with structured covariance matrices. ArXiv Preprint ArXiv:1607.06929.

SAID, S., BOMBRUN, L., BERTHOUMIEU, Y. and MANTON, J. H. (2017). Riemannian Gaussian distributions on the space of symmetric positive definite matrices. *IEEE Trans. Inform. Theory* **63** 2153–2170. MR3626862

SALEHIAN, H., CHAKRABORTY, R., OFORI, E., VAILLANCOURT, D. and VEMURI, B. C. (2015). An efficient recursive estimator of the Fréchet mean on a hypersphere with applications to medical image analysis. In *Mathematical Foundations of Computational Anatomy*.

SCHULDT, C., LAPTEV, I. and CAPUTO, B. (2004). Recognizing human actions: A local SVM approach. In *Pattern Recognition*, 2004. *ICPR* 2004. *Proceedings of the 17th International Conference on* **3** 32–36.

SRIVASTAVA, A., JERMYN, I. and JOSHI, S. (2007). Riemannian analysis of probability density functions with applications in vision. In *CVPR* 1–8.

STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds*, *Graphs*, *and Metric Spaces* (*Paris*, 2002). *Contemp*. *Math*. **338** 357–390. Amer. Math. Soc., Providence, RI. MR2039961

TUCH, D. S., REESE, T. G., WIEGELL, M. R. and WEDEEN, V. J. (2003). Diffusion MRI of complex neural architecture. *Neuron* **40** 885–895.

TURAGA, P., VEERARAGHAVAN, A. and CHELLAPPA, R. (2008). Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision. In *Computer Vision and Pattern Recognition*, 2008. *CVPR* 2008. *IEEE Conference on* 1–8. IEEE, New York.

WONG, Y. (1968). Sectional curvatures of Grassmann manifolds. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **60** 75–79. MR0229173

ZHANG, M. and FLETCHER, P. T. (2013). Probabilistic principal geodesic analysis. In *Advances in Neural Information Processing Systems* 1178–1186.

ZILLER, W. (2007). Examples of Riemannian manifolds with non-negative sectional curvature. In *Surveys in Differential Geometry. Vol*. *XI*. *Surv. Differ. Geom*. **11** 63–102. Int. Press, Somerville, MA. MR2408264

ZIMMERMANN, R. (2017). A matrix-algebraic algorithm for the Riemannian logarithm on the Stiefel manifold under the canonical metric. *SIAM J*. *Matrix Anal*. *Appl*. **38** 322–342. MR3639577

DEPARTMENT OF CISE
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611
USA
E-MAIL: rchakraborty@ufl.edu
vemuri@ufl.edu