# A geometric framework for statistical analysis of trajectories with distinct temporal spans*

Rudrasis Chakraborty[1], Vikas Singh[2], Nagesh Adluru[2], and Baba C. Vemuri[1]
[1]Department of CISE, University of Florida
[2]University of Wisconsin–Madison
https://youtu.be/PeID4r9SSxM

## Abstract

*Analyzing data representing multifarious trajectories is central to the many fields in Science and Engineering; for example, trajectories representing a tennis serve, a gymnast's parallel bar routine, progression/remission of disease and so on. We present a novel geometric algorithm for performing statistical analysis of trajectories with distinct number of samples representing longitudinal (or temporal) data. A key feature of our proposal is that unlike existing schemes, our model is deployable in regimes where each participant provides a* different *number of acquisitions (trajectories have different number of sample points or temporal span). To achieve this, we develop a novel method involving the parallel transport of the tangent vectors along each given trajectory to the starting point of the respective trajectories and then use the span of the matrix whose columns consist of these vectors, to construct a linear subspace in $\mathbf{R}^m$. We then map these linear subspaces (possibly of distinct dimensions) of $\mathbf{R}^m$ on to a single high dimensional hypersphere. This enables computing group statistics over trajectories by instead performing statistics on the hypersphere (equipped with a simpler geometry). Given a point on the hypersphere representing a trajectory, we also provide a "reverse mapping" algorithm to uniquely (under certain assumptions) reconstruct the subspace that corresponds to this point. Finally, by using existing algorithms for recursive Fréchet mean and exact principal geodesic analysis on the hypersphere, we present several experiments on synthetic and real (vision and medical) data sets showing how group testing on such diversely sampled longitudinal data is possible by analyzing the reconstructed data in the subspace spanned by the first few principal components.*

*We thank Anuj Srivastava (FSU) for feedback on a preliminary version of the paper and Yunyang Xiong (Wisconsin) for carefully curating the Gymnastics dataset used here. This research was funded in part by the NSF grant IIS-1525431 and IIS-1724174 (RC, BCV), Waisman IDDRC U54-HD090256 (NA), NSF CAREER award 1252725 (VS), and NIH grants AI117924/NIH CPCP and R01EB022883 (NA, VS).

## 1. Introduction

In many fields of science and engineering, one encounters data in the form of trajectories i.e., a one parameter family of multi-variate data, where the parameter describing the family is most commonly time or scale but can be any other parameter pertinent to the application. In Computer Vision (specifically in sports vision), a common example is the actions of an athlete such as the serve of a tennis player, a gymnast's routine, a golfer's swing and so on. In medical applications, analyzing a time course of structural or functional images to assess progress or remission of a disease in response to treatment is central to effective diagnosis. There is abundant literature on longitudinal (time course) data analysis where features of choice are *scalar* and/or *vector-valued*. However, with the advent of high throughput computing resources, applications are increasingly using sophisticated and rich feature sets such as manifold-valued features that are capable of capturing much more information contained in the raw imaging data than, say, scalar and vector-valued features.

There is a rigorous body of work in vision demonstrating how leveraging the *structure* (or geometry) of the data can yield advantages. For example in medical applications, papers in the mid-1990s already showed that the analysis of *shapes* [21] improved our ability to quantify a disease-specific signal, not otherwise possible. The interface between geometry/structure and analysis methods has offered effective practical tools — for instance, in Medical imaging applications, analysis of diffusion weighted Magnetic Resonance images where manifold-valued features such as the diffusion tensors that are symmetric positive definite (SPD) matrices that capture the diffusional behavior of water molecules at each image voxel may be inferred from the raw diffusion MRI data. Motivated by other applications, we have extensions of standard statistical machine learn-

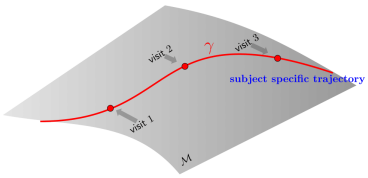
ing tools to the unit Hilbert sphere, probability distributions and other "structured" objects, i.e., where the samples are drawn from a space which satisfies a manifold characterization (e.g., covariance matrices). Algorithms for regression [28], principal components analysis, dictionary learning and others are readily available. Unfortunately, few such differential geometry inspired algorithms for image analysis exist for the *longitudinal* regime where outside of [35, 17, 40, 27, 25], the literature remains sparse. All such methods, however, cannot cope with trajectories of *distinct* number of samples within a group or across groups that are commonly encountered in group-wise longitudinal data analysis problems. In Computer Vision, several researchers have exploited manifold valued features such as covariance descriptors [44, 41, 11, 45, 37], image sets as linear subspaces [43, 8, 31, 22, 39, 33] and many others. Several of these proposals have dealt with analysis of videos for gait analysis, action recognition, dynamic textures and so on. For a comprehensive survey of linear dynamical system based modeling to analyze videos for a various such tasks, we refer the reader to [4]. These techniques do not address the question of statistical group-wise analysis of manifold-valued trajectories, each with a different number of samples.

**Goals.** Consider a setting, common across many longitudinal imaging studies or temporal data analysis tasks. We "track" a participant where at each visit, we acquire a manifold valued measurement (feature). This may be a shape (a sample on a shape manifold) or a diffusion tensor image which is a sample from a product space of an SPD manifold ("product" pertains to the number of voxels in an image). Of course, if every subject provided $p$ measurements each, we can repurpose existing algorithms for this task. The difficulty is that in general, due to logistic or financial reasons, *the number of samples from each subject are different*. When a subject joins the study late (or drops out), we get left (or right) censored data; in other cases, some intermediate visits may be missing. Imputation schemes are limited for manifold-valued data; so a practitioner is faced with two poor choices: (a) neglect the geometry of the space and shoehorn off the shelf techniques (problematic, both theoretically and empirically) or (b) only include participants with a full set of acquisitions (reduced sample sizes and corresponding decrease in statistical power). What is needed are frameworks that enable operating on "incomplete" longitudinal manifold-valued measures where incomplete refers to the nuisance of different number of samples/visits (temporal span) for each subject.

**Contributions.** This paper presents a novel algorithm to perform statistical analysis on the space of trajectories of manifold-valued measurements. A trajectory is a "path" on a Riemannian manifold comprised of a set of longitudinally acquired samples (points on the manifold). A salient feature of our technique is that trajectories with *different* number of samples are allowed, i.e., the number of points on a trajectory is *not* assumed to be a constant across the cohort (subjects). Our method involves parallel transporting the tangent vectors along each given trajectory (not necessarily a geodesic on the known data manifold) to the starting point of the respective given trajectories and then using the span of the matrix whose columns consist of these vectors, to construct a linear subspace of distinct dimension in $\mathbf{R}^m$. Then, using a result [13], we propose an algorithm to embed each linear subspace of distinct dimension (corresponding to a trajectory) into a single hypersphere. The hypersphere has a simple geometry which makes it more amenable than other alternatives [35, 25] to compute statistics. We also provide a procedure to identify the subspace which corresponds to a given point on the hypersphere. Within various settings (e.g., on OASIS data, Human Connectome project data, action recognition), we show the utility of this algorithm. Our results show that manifold-valued longitudinal datasets with different number of samples per subject can be easily handled within a flexible and efficient procedure. Further, our technique does not make the (restrictive) assumption that the given trajectory is a geodesic on the data manifold. This formulation and its analysis is the **main contribution** of this work.

## 2. Our proposed algorithms and analysis

**Preliminaries.** We first define the space of trajectories and then present the theory to compute statistics on this space. We define a trajectory $\gamma$ (see inline figure) to be a path that consists of a set of $p$ points on a Riemannian manifold $\mathcal{M}^m$ of dimension $m$ (inline figure shows $p = 3$).




Let $\{\gamma_i\}_{i=1}^N$ be a set of $N$ trajectories on $\mathcal{M}$, where $\gamma_i$ has $p_i$ sample points (note that this allows for trajectories with different number of sample points). Further, as each trajectory has a 'time" ordering (or an ordering with respect to any other variable), we can order the data point for $\gamma_i$ as $X_1^i, \cdots X_{p_i}^i$. To facilitate presentation of our theoretical results, we make the following mild **assumptions** about the trajectories:

**Assumption.** ***1)*** *For each trajectory $\gamma_i$, the sequence of $p_i$ data points lie on a continuous curve in $\mathcal{M}$.*

***2)*** *Without any loss of generality, we assume that $X_1^i$ is the starting point of $\gamma^i$, for all $i$.*

***3)*** *$\left\{X_1^i\right\}_{i=1}^N$ lie within a "regular geodesic ball" of radius $\pi/(2\sqrt{\kappa})$, $\kappa$ is the sectional curvature (we refer the readers to [26] for definition of regular geodesic ball). This assumption ensures that Fréchet mean (FM) [20] of $\left\{X_1^i\right\}_{i=1}^N$ exists and is unique [3].*

*4) For each trajectory, two consecutive data points can be joined by a geodesic, i.e., the Riemannian inverse exponential map [16] between two consecutive data points exists.*

*5) We assume $p_i < m$ for all $i$ (usually, $p_i \ll m$).*

**Grassmanian.** We will use $\text{Gr}(r, m)$ to denote the Grassmannian, i.e., the manifold of $r < m$ dimensional subspaces in $\mathbf{R}^m$ and $\text{Col}(\cdot)$ gives the column span operator, i.e., $\text{Col}(A)$ returns the linear subspace spanned by the columns of matrix $A$.

**Definition 1.** *Given a trajectory $\gamma$, i.e., a set of $p$ points, $X_1, \cdots X_p$ in $\mathcal{M}^m$, we compute $T_{X_j}\mathcal{M} \ni v_j = Log_{X_j}(X_{j+1})$, where $j = 1, \cdots, p-1$. Log is the inverse exponential map (Log exists because of Assumption 4). Thus, we identify $\gamma$ with a point in the product space $\underbrace{T\mathcal{M} \times \cdots \times T\mathcal{M}}_{(p-1) \text{ times}}$ via $\gamma \mapsto (X_1, v_1, \cdots v_{p-1})$, where $T\mathcal{M}$ is the tangent bundle of $\mathcal{M}$.*

The above identification has the following **properties**:

**Properties.** *1. The above identification is well-defined and is a bijection (this is trivial to show).*

*2. Since the tangent space of an $m$-dimensional manifold is isomorphic to $\mathbf{R}^m$, hence, in the above definition, $\underbrace{T\mathcal{M} \times \cdots \times T\mathcal{M}}_{(p-1) \text{ times}} \cong \mathcal{M} \times (\mathbf{R}^m)^{\oplus(p-1)}$, where $\oplus$ is the direct sum. Notice that since each $T_{X_j}\mathcal{M}$ has a different anchor point (since the base point $X_j$ varies), one needs to treat $v_i$ and $v_j$ as vectors in $TX_i\mathcal{M}$ and $TX_j\mathcal{M}$ respectively, but not as vectors in $\mathbf{R}^m$.*

*3. If $\mathcal{M}$ is parallelizable [16], the isomorphism in the property above is a diffeomorphism, e.g., since all Lie groups [16] are parallelizable, and if $\mathcal{M}$ is a Lie group, the above identification is a diffeomorphism.*

*4. If the manifold $\mathcal{M}$ is translated, let $\widetilde{\gamma}$ be the translated $\gamma$. Then, by the above identification, $\widetilde{v}_j = v_j$, for all $j$, and $\widetilde{X}_1$ and $X_1$ will be related by the translation, i.e., the above identification is translation invariant.*

**Setting up the space of trajectories.** Now, we parallel transport each $v_j$ from $T_{X_j}\mathcal{M}$ to $T_{X_1}\mathcal{M}$, for all $j = 2, \cdots p-1$. With a slight abuse of notation, we denote the parallel transported vectors by $\{v_j\}$. Since after the parallel transport operation, all $v_j$s lie in $T_{X_1}\mathcal{M}$, i.e., lie in the same vector space (which is isomorphic to $\mathbf{R}^m$), we form a matrix $V$ of dimension $m \times (p-1)$ whose $j^{th}$ column is $\iota(v_j)$, where $\iota : T_{X_1}\mathcal{M} \to \mathbf{R}^m$ is an isomorphism. Since we will be working with the matrix $V$, for the sake of notational simplicity, we will use $v_j$ to denote $\iota(v_j)$.

Let $\mathcal{V}$ be the column span of $V$, i.e., $\mathcal{V} = \text{Col}(V)$, then $\mathcal{V} \in \text{Gr}(r, m)$, where $r \leq (p-1)$ is the rank of $V$. Hence, using the identification of $\gamma$ in Definition 1, we can identify the space of trajectories with the product space of $\mathcal{M} \times \text{Gr}(r^\gamma, m)$, where $r^\gamma$ is the rank of $V$. Moreover, observe that $r^\gamma$ may be different for different trajectories $\gamma$. In other words, different trajectories correspond to different dimensional subspaces in $\mathbf{R}^m$. We should point out that although $r^\gamma$ can be different for different trajectories $\gamma$, they are all still subspaces in $\mathbf{R}^m$, as all the trajectories are on $\mathcal{M}$ (of dimension $m$).

**Remarks about this representation.** Note that our representation of trajectories is very general and unlike the previous methods, does not require that each trajectory should be a geodesic path [25, 35], or consists of an arbitrarily fixed number of points [40]. Also, when each trajectory has 2 points, our identification is same as in [25, 35] (as a topological space, not as a manifold as we use a different metric), i.e., our formulation is a generalization of [25, 35]. Moreover, by the above identification, we do not require linear independence of the points on a trajectory. This is a desirable property since, in many medical imaging problems, where a sequence of scans of a subject are often acquired longitudinally, the independence assumption is violated. *To the best of our knowledge, this is the first paper dealing with such a general setting for statistical analysis on the space of trajectories*.

**Ingredients for setting up a mapping.** As each trajectory may end up residing on a product of $\mathcal{M}$ and a Grassmann manifold of distinct dimension (recall that $r^\gamma$ may vary based on trajectory $\gamma$), we now propose a way to map each $\mathcal{V}^\gamma$ (note that the identification of trajectory $\gamma$ is $(X_1^\gamma, \mathcal{V}^\gamma) \in \mathcal{M} \times \text{Gr}(r^\gamma, m)$) onto a hypersphere. Given $\mathcal{V}^\gamma \in \text{Gr}(r^\gamma, m)$, the projection matrix, $P_\mathcal{V}^\gamma$ onto $\mathcal{V}^\gamma$ is defined by [12]:

$$P_\mathcal{V}^\gamma = \widetilde{V}\left(\widetilde{V}^T\widetilde{V}\right)^{-1}\widetilde{V}^T, \tag{1}$$

where $\widetilde{V}$ is a basis of $\mathcal{V}^\gamma$. Note that $P_\mathcal{V}^\gamma$ is a well-defined identification of $\mathcal{V}^\gamma$ as $P_\mathcal{V}^\gamma$ is independent of the basis of $\mathcal{V}^\gamma$, due to the following Lemma (stated without proof).

**Lemma 1.** *$P_\mathcal{V}^\gamma$ is independent of the choice of the basis of $\mathcal{V}^\gamma$ [12].*

## 2.1. Trajectories with distinct number of samples (temporal span)

Now, we state some properties of $P_\mathcal{V}$ (we drop the superscript for simplicity), which will be used in subsequent sections ([12] includes more details about these properties).

Fact **1)** $P_\mathcal{V}$ is a symmetric positive semi-definite matrix.

Fact **2)** $P_\mathcal{V}$ is an idempotent matrix, i.e., $P_\mathcal{V}^2 = P_\mathcal{V}$, its eigen values are either 0 or 1.

Fact **3)** The Frobenius norm of $P_\mathcal{V}$ is $\sqrt{r}$ where $\mathcal{V}$ is an $r$-dimensional subspace of $\mathbf{R}^m$.

Fact **4)** The rank of $P_\mathcal{V}$ is $r$.

Fact 3 above gives us a way to map $P_{\mathcal{V}}$ (after vectorization) onto $\mathbf{S}^{m^2-1}$ of radius $\sqrt{r}$. This is an isometric embedding as was shown in [13]. So, given $\mathcal{V}^{\gamma_1}$ and $\mathcal{V}^{\gamma_2}$ (the matrices constructed from the parallel transported tangent vectors corresponding to trajectories $\gamma_1$ and $\gamma_2$ respectively), we can map them on $\mathbf{S}^{m^2-1}$ of radius $\sqrt{r^{\gamma_1}}$ and $\sqrt{r^{\gamma_2}}$ respectively. Now, we can scale $P_{\mathcal{V}}^{\gamma_1}$ and $P_{\mathcal{V}}^{\gamma_2}$ by $\sqrt{r^{\gamma_1}}$ and $\sqrt{r^{\gamma_2}}$ respectively to map them both onto the unit hypersphere, $\mathbf{S}^{m^2-1}$. The question we may ask is: *Is this an injective mapping, i.e., given two different subspaces, $\mathcal{V}^{\gamma_1}$ and $\mathcal{V}^{\gamma_2}$ with the same starting point, i.e., $X_1^{\gamma_1} = X_1^{\gamma_2}$, can they map onto the same point on $\mathbf{S}^{m^2-1}$?* If the two subspaces are not subspaces of each other, the answer is no. This assumption is satisfied quite commonly in practice and we will make this assumption here as well. Now, we will formally state and prove the following theorem which shows that the above mapping is injective.

**Theorem 1.** *Let $\mathcal{V}^{\gamma_1}$ and $\mathcal{V}^{\gamma_2}$ be two linear subspaces on $\mathbf{R}^m$. Without any loss of generality, assume $r^{\gamma_1} \leq r^{\gamma_2}$. Assume $\mathcal{V}^{\gamma_1}$ is not a subspace of $\mathcal{V}^{\gamma_2}$. Then, the above mapping is injective.*

*Proof.* Let us assume that the mapping is not injective, i.e., $\text{vec}(P_{\mathcal{V}}^{\gamma_1})/\sqrt{r^{\gamma_1}} = \text{vec}(P_{\mathcal{V}}^{\gamma_2})/\sqrt{r^{\gamma_2}}$. Then, $P_{\mathcal{V}}^{\gamma_1} = cP_{\mathcal{V}}^{\gamma_2}$, where $c = \sqrt{\frac{r^{\gamma_1}}{r^{\gamma_2}}}$. Observe that, given $P_{\mathcal{V}}^{\gamma_1}$, the corresponding $\mathcal{V}^{\gamma_1} = \text{Col}(\widetilde{U})$, where $U\Sigma U^T = \text{eig}(P_{\mathcal{V}}^{\gamma_1})$ and $\widetilde{U}$ is the first $r^{\gamma_1}$ columns of $U$. Now since, $P_{\mathcal{V}}^{\gamma_1} = cP_{\mathcal{V}}^{\gamma_2}$, their eigen decompositions are the same, i.e., $\mathcal{V}^{\gamma_1}$ is a subspace of $\mathcal{V}^{\gamma_2}$ (which is a contradiction to the assumption). □

We will now present the forward mapping algorithm in Alg.-1 to map a trajectory $\gamma$ onto the product space of $\mathcal{M}$ and the unit hypersphere.

**Algorithm 1:** Algorithm to map a trajectory onto the $\mathcal{M} \times \mathbf{S}^{m^2-1}$.

**Input**: $\gamma$ consists of $p$ points on $\mathcal{M}$ ($M$ is of dimension $m$)
**Output**: $(X_1, \boldsymbol{s}^{\gamma}) \in \mathcal{M} \times \mathbf{S}^{m^2-1}$

1. Let the starting point of $\gamma$ be $X_1$;
2. Compute tangent vector $v_j$ from $X_j$ to $X_{j+1}$, $j = 1, \cdots p-1$;
3. Parallel transport all the vectors to $T_{X_1}\mathcal{M}$ and column stack them to form a matrix $\widetilde{V}$ of dimension $m \times (p-1)$;
4. Orthonormalize $\widetilde{V}$ using the Gram-Schmidt orthonormalization to get $V$, let the rank of $V$ be $r^{\gamma}$ ;
5. Compute the projection matrix $P_{\mathcal{V}}^{\gamma}$ using Eq. 1. ;
6. Compute $\boldsymbol{s}^{\gamma} = \text{vec}(P_{\mathcal{V}}^{\gamma})/\sqrt{r^{\gamma}}$.

In Fig. 1, we give a pictorial description of our proposed framework. Equipped with the algorithm to map from the space of trajectories to $\mathcal{M} \times \mathbf{S}^{m^2-1}$, we will now conduct statistical analysis on the product space, which has a simpler geometry (relative to the space of trajectories). We will first define a Gaussian distribution on the hypersphere, $\mathbf{S}^n$. It is well-known in differential geometry that $\mathbf{S}^n$ is a homogeneous space and can be identified with $O(n+1)/O(n)$

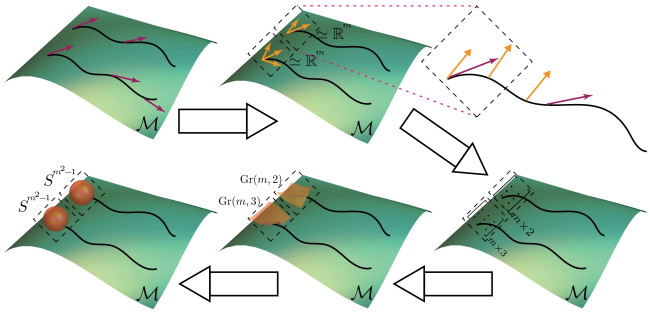



Figure 1: The pictorial description of the framework to map trajectories with different number of sample points.

where $O(n)$ is the compact Lie group of orthogonal matrices [23]. Now, we will briefly give the geometry of a homogeneous space $\mathcal{N}$. For a good reference on homogeneous spaces, we refer the reader to [23].

## 2.2. Defining distributions on a homogeneous space

In this section, we will briefly summarize the results from [10] relating to rigorously defining a Gaussian distribution on a homogeneous space and then specialize it to $\mathbf{S}^n$ (which is identified with a homogeneous space). First, we will summarize the differential geometry of a homogeneous space $\mathcal{N}$, which is needed as background material.

Let $(\mathcal{N}, \mathfrak{g})$ be a Riemannian manifold with a Riemannian metric $\mathfrak{g}$. Let $d$ be the metric induced by the Riemannian metric $\mathfrak{g}$. Let $G$ be the set of all isometries of $\mathcal{N}$, i.e., given $g \in G$, $d(g.X, g.Y) = d(X, Y)$, for all $X, Y \in \mathcal{N}$. Let $O \in \mathcal{N}$ and let $H = \text{Stab}(O) = \{h \in G | h.O = O\}$ (Stab is abbreviation for Stabilizer). We say $G$ acts *transitively* on $\mathcal{N}$, *iff* given $X, Y \in \mathcal{N}$, there exists a $g \in G$ such that $Y = g.X$.

**Definition 2.** *Let $N$ be a Riemannian manifold. Let $G = I(\mathcal{N})$ act transitively on $\mathcal{N}$ and $H = Stab(O)$, $O \in \mathcal{N}$ (called the "origin" of $\mathcal{N}$) is a subgroup of $G$. Then, $\mathcal{N}$ is a homogeneous space and can be identified with the quotient space $G/H$ under the diffeomorphic mapping $gH \mapsto g.O, g \in G$ [23].*

From the definition of a homogeneous space, we know that the Riemannian metric $\mathfrak{g}$ at $X$ is invariant under the group operation $X \mapsto g.X$, hence the volume element $d\nu$ is also preserved.

**The Gaussian distribution on a homogeneous space:** Let $M \in \mathcal{N}$ denote the location parameter and $\sigma > 0$ be the scale parameter. Now, we will define the Gaussian distribution function on a homogeneous space $\mathcal{N}$ with respect to an appropriately defined probability measure $d\nu$ [36] as:

$$f_X(M, \sigma) = \frac{1}{Z(\sigma)} \exp\left(-\frac{d^2(X, M)}{2\sigma^2}\right) \tag{2}$$

**Theorem 2.** *The normalization factor in Eq. 2 i.e., $Z(M,\sigma) = \int f_X(M,\sigma)\, d\nu(X)$ is a constant and is independent of $M$, i.e., the function in Eq. 2 is a valid probability density function (see [10] for a proof).*

**Algorithm to reconstruct a trajectory:** With a Gaussian distribution on homogeneous space (and in turn on $S^{m^2-1}$) defined as above, we now give an algorithm to reconstruct trajectory from a given sample in $\mathcal{M} \times \mathbf{S}^{m^2-1}$ (*Note that under the condition in Theorem 1 this is a well-defined mapping*). We will assume that $\mathcal{M}$ is a matrix Lie group. This means that $\mathcal{M}$ is a group and is a smooth manifold such that the group operations (multiplication and inverses) are smooth maps. Let $e$ be the identity element of Lie group $\mathcal{M}$. The tangent space at $e$, i.e., $T_e\mathcal{M}$ is called the Lie algebra $\mathfrak{m}$ corresponding to $\mathcal{M}$. Let $U, V \in \mathfrak{m}$, then the left-invariant metric, $g$ on $\mathfrak{m}$ is defined as $g(U,V) = \text{trace}(U^T V)$, i.e., $\mathfrak{m}$ is equipped with an Euclidean metric. Moreover, given $U \in \mathfrak{m}$, one can parallel transport $U$ from $\mathfrak{m}$ to $T_X\mathcal{M}$ by $\mathfrak{m} \ni U \mapsto XU \in T_X\mathcal{M}$. The Riemannian exponential map is defined as $\text{Exp}_X(U) = X\text{Exp}(X^{-1}U)$, where $U \in T_X\mathcal{M}$ and Exp is the matrix exponential. The Riemannian inverse exponential map is defined as $\text{Log}_X(Y) = X\text{Log}(X^{-1}Y)$, where Log is the matrix logarithm. We refer the reader to [23] for a good reference on Lie groups.

Now, using the Gaussian distribution defined on Lie groups in [19], and the Gaussian distribution on a homogeneous space (defined earlier), we can get a sample on $\mathcal{M} \times \mathbf{S}^{m^2-1}$, when $\mathcal{M}$ is a Lie group. We are now ready to develop an algorithm to obtain $\gamma$ from a point on $\mathcal{M} \times \mathbf{S}^{m^2-1}$.

**Algorithm 2:** Algorithm to recover a trajectory corresponding to a point in $\mathcal{M} \times \mathbf{S}^{m^2-1}$.

**Input**: $(X_1, \boldsymbol{s}^{\boldsymbol{\gamma}}) \in \mathcal{M} \times \mathbf{S}^{m^2-1}$ where $\mathcal{M}$ is a Lie group and $\boldsymbol{s}^{\boldsymbol{\gamma}}$ is a vectorized projection matrix
**Output**: $\gamma$ consists of $r$ points on $\mathcal{M}$
1 Arrange $\boldsymbol{s}^{\boldsymbol{\gamma}}$ in a $m \times m$ matrix $Y$ ;
2 Compute the rank of $Y$, let the rank be $r-1$ ;
3 Perform eigen decomposition of $Y$, i.e., $\widetilde{V}\Sigma\widetilde{V}^T = Y$, then, assign $V$ to be the first $r-1$ columns of $\widetilde{V}$. Note that in order for $Y$ to be the projection matrix of $\mathcal{V}$ (using Eq. 1), we assume that each column of $V$, i.e., $v_j$ lies on $\mathfrak{m}$, so that the metric is the Euclidean inner product;
4 Use $X_j v_j \in T_{X_j}\mathcal{M}$ to construct $X_{j+1}$ from $X_j$ (using the parallel transport of $v_j$ from $T_e\mathcal{M}$ to $T_{X_j}\mathcal{M}$), $j = 1, \cdots r-1$;
5 Return $\gamma$ consisting of $X_1, X_2, \cdots X_r$

We should point out that Alg.-2 assumes that the input is a vectorized projection matrix. But, any point on hypersphere may not be a vectorization of a projection matrix. We now give a projection algorithm which takes an arbitrary point on $\mathbf{S}^{m^2-1}$ and returns its closest point on $\mathbf{S}^{m^2-1}$ that has a preimage on $\text{Gr}(.,m)$. Note that, Alg.-2 can be applied to this closest point. Algorithm 3 is a projection algorithm from a square matrix to its closest symmetric positive semi-definite idempotent matrix. One can prove that this algorithm returns the closest projection matrix by an argument similar to Theorem 2 in [42]. Now, we will give expressions for the Riemannian exponential (denoted by Exp) and inverse exponential (denoted by Log) maps which will be required throughout the rest of the paper. Given $\mathbf{x}, \mathbf{y} \in \mathbf{S}^n$, the geodesic distance between $\mathbf{x}$ and $\mathbf{y}$, denoted by $d(\mathbf{x},\mathbf{y}) = \arccos(\mathbf{x}^t\mathbf{y})$. The exponential map at $\mathbf{x}$ is given by $\text{Exp}_{\mathbf{x}}(\mathbf{v}) = \cos(\|\mathbf{v}\|)\mathbf{x} + \sin(\|\mathbf{v}\|)\frac{v}{\mathbf{v}}$, where $\mathbf{v} \in T_{\mathbf{x}}\mathbf{S}^n$. The inverse exponential map between $\mathbf{x}$ and $\mathbf{y}$ as $\text{Log}_{\mathbf{x}}(\mathbf{y}) = \frac{\theta}{\sin(\theta)}(\mathbf{y} - \mathbf{x}\cos(\theta))$, where $\theta = d(\mathbf{x},\mathbf{y})$.

**Algorithm 3:** The projection algorithm.

**Input**: $\boldsymbol{s} \in \mathbf{S}^{m^2-1}$
**Output**: $\boldsymbol{s}^{\boldsymbol{\gamma}} \in \mathbf{S}^{m^2-1}$ which is an input of Alg.-2
1 Arrange $\boldsymbol{s}$ in a $m \times m$ matrix $Y$ ;
2 Compute the rank of $Y$, let the rank be $r$ ;
3 Perform the eigen decomposition of $Y$, i.e., $\widetilde{V}\Sigma\widetilde{V}^T = Y$, then, assign $V$ to be the first $r$ columns of $\widetilde{V}$ ;
4 Compute $P = VV^T$ ;
5 Vectorize $P$ and divide by $\sqrt{r}$ to get $\boldsymbol{s}^{\boldsymbol{\gamma}}$.

**MLE of $M$:** Given $\{\mathbf{x}_i\}_{i=1}^N \subset \mathbf{S}^n$, the Fréchet mean (FM) [20], $\boldsymbol{\mu}$ is defined as $\mu = \arg\min_{\mathbf{z}\in\mathbf{S}^n} \sum_{i=1}^N d^2(\mathbf{z},\mathbf{x}_i)$. The existence and uniqueness of FM is guaranteed if the samples lie within a "regular geodesic ball" of radius $\pi/2$ [3] (we refer the readers to [26] for definition of regular geodesic ball). We will now state (proof is in the supplementary section) that maximum likelihood estimator (MLE) of $M$ defined above is the FM.

**Lemma 2.** *Given, $\{\mathbf{x}_i\}_{i=1}^N \subset \mathbf{S}^n$ i.i.d. samples drawn from the Gaussian distribution whose support is within a geodesic ball of radius $< \pi/2$, the MLE of $M$ (defined in Eq. 2) is the FM of $\{\mathbf{x}_i\}_{i=1}^N$.*

Note that, although Alg.-2 assumes $\mathcal{M}$ to be a Lie group, it is also applicable to other special manifolds, e.g., space of symmetric positive definite matrices (SPD) and the hypersphere. The reason for assuming the Lie group structure is two fold (i) On a Lie group, the tangent space at $e$, i.e., $\mathfrak{m}$ or the Lie algebra is equipped with Euclidean metric, hence using $XX^T$ to get projection matrix is meaningful on $\mathfrak{m}$. (ii) After getting tangent vectors on $\mathfrak{m}$, we can do simple matrix multiplication to transport $v_j$ into $T_{X_j}\mathcal{M}$ in the Alg.-2.

Now, we will show that both these properties are satisfied for the manifold of SPD matrices (with the $GL$-invariant metric) [24] and the hypersphere (with the arc-length metric). Let $\mathcal{M}$ be a space of $m \times m$ SPD matrices, we can define $GL$-invariant metric, $g_X$ on this manifold as $g_X(U,V) = \text{trace}\left(X^{-1}UX^{-1}V\right)$, where $U, V \in T_X\mathcal{M}$. So, if $X$ is the identity matrix, clearly, $g_X$ is the Euclidean inner product, hence the property (i) above is satisfied.

Moreover, as the $m \times m$ invertible matrix (as a general linear group, $GL(m)$) acts on $\mathcal{M}$, the parallel transport is just a group operation. Hence, Alg.-2 is applicable to the SPD-manifold with a $GL$-invariant metric.

Let us consider the hypersphere $\mathbf{S}^n$. Any vector $\mathbf{v} \in \mathbf{R}^n$ can be projected onto $T_{\mathbf{x}}\mathbf{S}^n$ by the following operation $\mathbf{v} \mapsto \mathbf{v} - (\mathbf{v}^t\mathbf{x})\mathbf{x}$. Moreover, the parallel transport on $\mathbf{S}^n$ is in an analytic form, hence we can apply Alg.-2 on $\mathbf{S}^n$. As a side note, we would like to point out that equipped with the Log-Euclidean metric, the SPD manifold has a Lie group structure as shown in [5] and $\mathbf{S}^0$, $\mathbf{S}^1$ and $\mathbf{S}^3$ are the only hyperspheres which are Lie groups.

In the experiments, we have assumed $\mathcal{M}$ to be either a hypersphere or an SPD-manifold. We will use the incremental/recursive FM computation algorithm proposed in [38] to compute FM of samples on $\mathbf{S}^{m^2-1}$ and on the SPD-manifold, we will use the recursive algorithm for FM computation proposed in [24]. Later, we will perform principal geodesic analysis (PGA) on the space of trajectories by using *exact*-PGA on $\mathbf{S}^{m^2-1}$, presented in [9]. Both of these methods are extremely efficient and the consistency of the incremental/recursive FM estimator was proved in [38, 24].

## 3. Experiments

In this section, we demonstrate the application of the framework to answer three important questions that arise in neuroimaging and vision applications. **(1)** Can principal geodesics (PGs) offer efficient representations in detecting group differences in longitudinal neuroimaging studies? **(2)** How robust is our framework to missing temporal data or temporal data with varying number of time points? **(3)** Do principal geodesics offer features that are independent of the temporal spans of videos? Before we dive into the experiments to evaluate these questions, we present experiments using synthetic data computing the Fréchet mean estimation of trajectories.

**FM computation of trajectories for synthetic data:** We randomly generate geodesics on $\mathbf{S}^2$. We show the mean trajectory for these synthetic experiments in Fig. 2. We compared the results with [25], and as expected since all the trajectories are geodesic paths, our proposed method yields similar mean trajectory as that from the method in [25]. This serves as a sanity check showing that for simulated data, our results are consistent with an existing method from the literature.

**Efficient representation:** We use OASIS data [2] to demonstrate that using our framework, we can use PG to detect class/group structure. OASIS data [2] contains MR scans of demented, non-demented and "converted" patients. Patients who are labelled as "converted" are those who progressed from non-demented to demented during the study. This dataset contains at least two MR brain scans of 150 subjects, aged between 60 to 96 years old. For each patient, scans are separated by at least one year. The dataset includes patients of both sexes. In order to avoid gender effects, we use MR scans of female patients from 2-5 visits, which results in a dataset containing MR scans of 11 subjects with dementia (denoted by the letter 'D') and 12 subjects without dementia (denoted by 'ND') and 7 subjects of "converted" (denoted by the letter 'C') group. We first compute an atlas (using the method in [6]) from the MR scans of patients without dementia. After rigidly registering each MR scans to the atlas, we only consider intensity values in a prespecified region of interest (ROI), namely the corpus callosum (CC) that is known to be effected most by the disease process, from each image. Then, using the Scrödinger distance transform (SDT) [15] applied to the ROI, we map the CC shape to point on $\mathbf{S}^{2332}$. For each subject, we have 2-5 time points, i.e., the trajectories constitue varying # of time points.

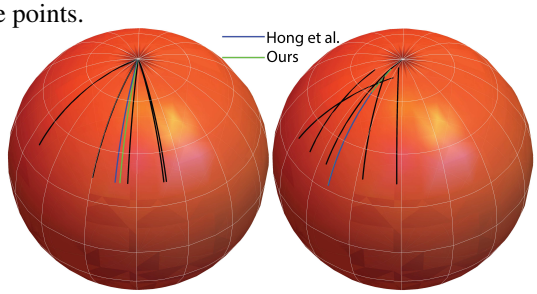



Figure 2: Trajectories are shown in black and the mean trajectories (using proposed method and method in [25]) are shown in green and blue respectively. The results from both methods are similar suggesting that our representation is reasonable.

We performed Principal geodesic analysis (PGA) to evaluate classification accuracy and group differences on the OASIS data. We take the first 10 principal geodesics (PGs) and perform reconstruction of the data. On the reconstructed data, we perform a pairwise group testing as follows. We first choose two classes and compute the distance between the two mean trajectories (mean trajectory from each class). Then, we randomly permute the class labels 10000 times. We then count the fraction of the times the distance between two group means computed with these random permutations is larger than the distance on the data with with correct permutation (class labels). This gives an approximation of the $p$-value which is reported in Table 1a. Note that a smaller value signifies that there is indeed a class structure preserved in the reconstructed data. We can see from the table that our framework preserves better class structure in the reconstructed data using the first 10 PGs since, the $p$-value is significantly smaller than that of [25]. Next, we will perform a pairwise leave-one-out classification with the PGs to see if our framework indeed gives better classification accuracy.

We use a linear SVM classifier on the PGs and report sensitivity (denoted by 'sn'), specificity (denoted by 'sp') and classification accuracy (denoted by 'ac') in Table 2. It

| Class names | Our Method | [25] |
|---|---|---|
| C vs. D | **0.065** | 0.53 |
| C vs. ND | **0.035** | 0.38 |
| D vs. ND | **0.051** | 0.46 |

(a) Comparison of approximated $p$-values.

| | Our Method | [25] |
|---|---|---|
| $R^2$ statistics | **0.64** | 0.37 |
| $p$-value | **0.032** | 0.087 |

(b) Comparison of regression results

Table 1: Statistical analysis on OASIS data.

is clear from the table that we achieve a better classification accuracy than [25].

**Robustness to data with varying time points:** In this section, we will demonstrate the performance of our framework to do statistical analysis on temporal data with varying time points or on temporal data with missing entries. OASIS data already included varying # of time points. In this section, we also use data from the Human Connectome project (HCP) to extract trajectories with missing time points. Before going into the details, we briefly describe the HCP data and how we extract the trajectories.

| Class names | Our Method | | | [25] | | |
|---|---|---|---|---|---|---|
| | **sn** | **sp** | **ac (%)** | **sn** | **sp** | **ac (%)** |
| C vs. D | **0.86** | **0.91** | **88.89** | **0.86** | 0.82 | 83.33 |
| C vs. ND | **0.86** | **1.00** | **94.73** | 0.71 | 0.83 | 78.95 |
| D vs. ND | **0.91** | **0.92** | **91.31** | **0.91** | 0.83 | 86.96 |

Table 2: Classification on OASIS data.

All subjects in the main HCP cohort were scanned on a dedicated 3 Tesla (3T) scanner. We analyzed the high-quality curated diffusion MR imaging (dMRI) data made publicly available on over 840 healthy adults from the WU-Minn consortium [46]. We obtained diffusion tensor images (DTI) from the dMRI data by non-linear fitting of the tensors to the diffusion weighted ($b = 1000$ s/mm$^2$) images. These DTI images were spatially normalized using DTI-TK [48], a non-linear diffeomorphic registration and template estimation pipeline, that can directly operate on the diffusion tensors using a log-Euclidean framework. Seventeen major white matter pathways were obtained by registering the publicly available IIT white matter atlas [47] to the HCP template using the ANTS software [6]. We analyzed DTI data from the fornix and the cingulum bundle.

Now, from this data, we build the trajectories as follows. We first divide ages of the subjects into the following bins: $[22, 25]$, $[26, 29]$, $[30, 33]$ and $[34, .)$. Next, we sample 20 subjects from each bin, for all bins. The average of these 20 gives us a virtual subject who is tracked across the bins. This is a single trajectory sample with each point on the trajectory belonging to a product space of 228, $3 \times 3$ SPD matrices. We replicate this process 500 times to get 500 virtual subjects who are tracked across all bins. Then, we randomly choose 2-4 bins for each subject to simulate a situation where we have missing entries corresponding to some time points. For a pictorial depiction of the trajectory generation for the connectome data, see Fig. 3.

One of the major tools to do statistical analysis is to perform regression between a set of independent and dependent variables. Now, we will analyze performance of a regressor in the situation where the data has varying time points or has missing entries. We will compare the performance between our formulation and the formulation proposed in [25] on both OASIS and HCP data.

For OASIS data, an important question to ask is: *Is there any relationship between the structure of corpus callosum and age?*. Recently, in [7], the authors have shown that there is indeed a relationship. Motivated by this result, here we ask the following question: *Is there any relationship between the changes in the structure of the corpus callosum and age?* Further, for different patients, we measured the changes on varying number of time points. We use the manifold-valued kernel regression (MVKR) technique proposed in [7] as the non-linear regressor. In order to evaluate the performance, we chose the $R^2$ statistic on Riemannian manifolds as defined in [18] as a measure. An $R^2$ statistic value close to 1 implies better regression performance. The comparative results are reported in Table 1b which clearly suggest that the regressor gives better $R^2$ statistic using our framework. Moreover, as the regression relationship is complex, so the approximation of a trajectory by a geodesic is a probable reason behind the poor $R^2$ statistics value given by [25]. Now, we perform a $t$-test with 1000 independent runs to check the statistical significance of the regression result. We reject the null hypothesis "$H_0$ = mean of the unexplained variance is not less than the mean of the data variance" with a significance level of 0.05. From the $t$-test result we can see that our results are statistically significant.

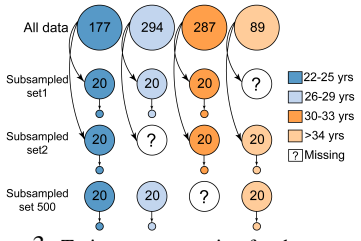



Figure 3: Trajectory generation for the connectome data (each bubble shows the number of samples).

Recall that in the HCP data, we have trajectories with missing entries. For this data, it is meaningful to ask how the behavioral measure of a person relates to the changes in the brain scans, i.e., on the virtual subject, as we track the changes of brain scans, can we predict the behavioral scores. We have two such scores namely, `ProcSpeed_Unadj` (denoted by 'pU') and `ListSort_Unadj` (denoted by 'lU'). These scores measure processing speed of subjects in sorting a list of items. As before, we perform kernel regression (MVKR) and compute the $R^2$ statistic. The comparative results are reported in Table 3. The results indicate a good $R^2$ statistic value using our method and an unsatisfactory performance by the method in [25]. As before, we also perform a $t$-test on 1000 independent runs to check how statistically significant is the $R^2$ statistic value. We choose the null hypothesis as in the case of OASIS data and reject

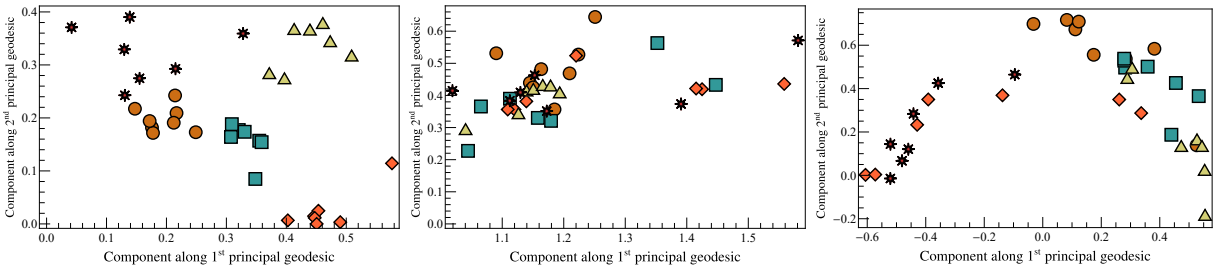



Figure 4: Summarization of the Gymnastics data (five gymnastic activities) using *Left:* Our method, *Middle:* [30] and *Right:* [25]. Note the clear class separation obtained by our method.

with significance level 0.05. Moreover, in contrast to the baseline, our method yields statistically significant results. We suspect that this is because in the HCP data, trajectories are "complicated", i.e., deviate from a geodesic; so, using a geodesic-based formulation may not be appropriate for regression on this data.

**Length (of time) invariant representation:** In computer vision, a common type of temporal data analysis is in calculating statistical summaries of video data. In this section, we deal with videos of gymnastic routines from 2012 London Olympics [1]. Moreover, this data is of varying temporal span, so it would be interesting to statistically summarize this data. Each video is of dimension of $640 \times 360$ and the frame rate is 30 fps. We collected videos of 5 Gymnastic activities, where each activity is performed by 8 gymnasts. We sampled this video using $1/3$ fps.

| Behavioral Scores | Our Method | | [25] | |
|---|---|---|---|---|
| | $R^2$ statistics | $p$-value | $R^2$ statistics | $p$-value |
| $p$U | **0.78** | **0.017** | 0.02 | 0.83 |
| $l$U | **0.75** | **0.021** | 0.16 | 0.75 |

Table 3: Regression results on HCP data.

From each frame, we extracted HOG features [14], using the following parameter values: *B*locksize $= 2$, Cellsize $= 16$, Blockoverlap $= 4$, Number of Bins $= 9$. We normalize the HOG features to map it to $\mathbf{S}^{1763}$. We construct the trajectory from each gymnast's video by taking each frame as a point on the trajectory. Due to the varying time span of the videos, we get trajectories of varying number of time points. A sample trajectory from each act is shown in Fig. 5.

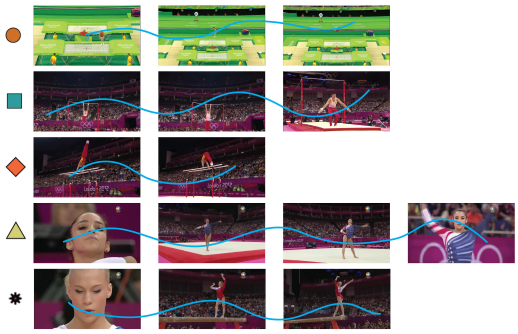


Figure 5: Sample trajectories and corresponding legends used in Fig. 4.

We report results of groupwise statistical summarization of the Gymnastics routines from several gymnasts across the world. The summarization is depicted in the form of a biplot showing ability of a method employed to group gymnasts within groups. In this experiment, we performed PGA on the trajectories representing gymnast routines, using our formulation. We summarize the data, in $\mathbf{R}^2$ by taking the component along the first two PGs. In addition to a comparison with [25], we also compared results from using space-time features. We first used Harris3D detector [30] to extract spatio-temporal interest point from each video. Then from each interest point, we calculate HOG and HOF features [32]. We use the implementation available on-line [29] with standard parameter settings. Then, we use kernel-PCA [34] (with a Gaussian kernel) on the feature vector to get the first two PCs. The comparison is depicted in Fig. 4, where, we can see that our formulation yields the best summary in terms of preserving better structure within the same activity.

## 4. Conclusions

We presented a novel geometric framework and algorithms for computing statistics on the space of trajectories representing longitudinal data. The salient features of our algorithm are: (i) it can seamlessly cope with trajectories of distinct temporal spans and (ii) the framework maps each trajectory of varying # of time points represented by a linear subspace of $\mathbf{R}^m$ on to a single finite dimensional hypersphere. Since, the geometry of the hypersphere is simple and yields analytic expressions for most geometric quantities of interest here, it gives our algorithm an edge over the competition. Finally, unlike most existing methods for trajectory modeling, our method does not require that all the sample points of a trajectory lie on a geodesic. We presented experiments demonstrating how group testing on longitudinal data with different number of time samples is possible by analyzing the reconstructed data in the subspace spanned by the first few PGs. We also presented experiments demonstrating robustness of our framework to missing time points. Finally, we performed a statistical summarization of temporal data of varying time spans and compared the performance with the state-of-the-art.

## References


[1] London 2012. https://www.olympic.org/london-2012.

[2] OASIS. http://www.oasis-brains.org/.

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

[4] B. Afsari and R. Vidal. Distances on spaces of high-dimensional linear stochastic processes: A survey. In *Geometric Theory of Information*, pages 219–242. Springer, 2014.

[5] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic resonance in medicine*, 56(2):411–421, 2006.

[6] B. B. Avants, N. Tustison, and G. Song. Advanced normalization tools (ANTS). *Insight J*, 2:1–35, 2009.

[7] M. Banerjee, R. Chakraborty, E. Ofori, M. S. Okun, D. E. Viallancourt, and B. C. Vemuri. A nonlinear regression technique for manifold valued data with applications to medical image analysis. In *CVPR*, pages 4424–4432, 2016.

[8] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto. Recognition of human gaits. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE, 2001.

[9] R. Chakraborty, D. Seo, and B. C. Vemuri. An efficient exact-pga algorithm for constant curvature manifolds. *arXiv preprint arXiv:1603.03984*, 2016.

[10] R. Chakraborty and B. Vemuri. Statistics on the (compact) stiefel manifold: Theory and applications. *http://arxiv.org/abs/1708.00045*, 2017.

[11] G. Cheng and B. C. Vemuri. A novel dynamic system in the space of spd matrices with applications to appearance tracking. *SIAM journal on imaging sciences*, 6(1):592–615, 2013.

[12] R. Christensen. *Plane answers to complex questions: the theory of linear models*. Springer Science & Business Media, 2011.

[13] J. H. Conway, R. H. Hardin, and N. J. Sloane. Packing lines, planes, etc.: Packings in grassmannian spaces. *Experimental mathematics*, 5(2):139–159, 1996.

[14] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[15] Y. Deng, A. Rangarajan, S. Eisenschenk, and B. C. Vemuri. A riemannian framework for matching point clouds represented by the schrödinger distance transform. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 3756–3761. IEEE, 2014.

[16] M. P. DoCarmo. *Riemannian Geometry*. Birkhauser, 1992.

[17] S. Durrleman, X. Pennec, A. Trouvé, J. Braga, G. Gerig, and N. Ayache. Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *IJCV*, 103(1):22–59, 2013.

[18] P. T. Fletcher. Geodesic regression and the theory of least squares on riemannian manifolds. *IJCV*, 105(2):171–185, 2013.

[19] P. T. Fletcher, S. Joshi, C. Lu, and S. M. Pizer. Gaussian distributions on lie groups and their application to statistical shape analysis. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 450–462. Springer, 2003.

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

[21] U. Grenander. *Elements of pattern theory*. JHU Press, 1996.

[22] J. Hamm and D. D. Lee. Grassmann discriminant analysis: a unifying view on subspace-based learning. In *Proceedings of the 25th international conference on Machine learning*, pages 376–383. ACM, 2008.

[23] S. Helgason. *Differential geometry, Lie groups, and symmetric spaces*, volume 80. Academic press, 1979.

[24] J. Ho, G. Cheng, H. Salehian, and B. C. Vemuri. Recursive karcher expectation estimators and geometric law of large numbers. In *AISTATS*, pages 325–332, 2013.

[25] Y. Hong, N. Singh, R. Kwitt, and M. Niethammer. Group testing for longitudinal data. In *IPMI*, pages 139–151, 2015.

[26] W. S. Kendall. Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proc. of LMS*, 3(2):371–406, 1990.

[27] H. Kim, N. Adluru, H. Suri, B. C. Vemuri, S. C. Johnson, and V. Singh. Riemannian nonlinear mixed effects models: Analyzing longitudinal deformations in neuroimaging. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[28] H. J. Kim, N. Adluru, M. D. Collins, M. K. Chung, B. B. Bendlin, S. C. Johnson, R. J. Davidson, and V. Singh. Multivariate general linear models (MGLM) on Riemannian manifolds with applications to statistical analysis of diffusion weighted images. In *CVPR*, Columbus, Ohio, June 2014.

[29] I. Laptev. https://www.di.ens.fr/~laptev/download.html#stip, 2005.

[30] I. Laptev. On space-time interest points. *International journal of computer vision*, 64(2-3):107–123, 2005.

[31] D. Lin, S. Yan, and X. Tang. Pursuing informative projection on grassmann manifold. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1727–1734. IEEE, 2006.

[32] B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. 1981.

[33] Y. M. Lui and J. R. Beveridge. Grassmann registration manifolds for face recognition. In *European Conference on Computer Vision*, pages 44–57. Springer, 2008.

[34] S. Mika, B. Schölkopf, A. J. Smola, K.-R. Müller, M. Scholz, and G. Rätsch. Kernel pca and de-noising in feature spaces. In *NIPS*, volume 11, pages 536–542, 1998.

[35] P. Muralidharan and P. T. Fletcher. Sasaki metrics for analysis of longitudinal data on manifolds. In *CVPR*, pages 1027–1034, 2012.

[36] X. Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *JMIV*, 25(1):127–154, 2006.

[37] F. Porikli, O. Tuzel, and P. Meer. Covariance tracking using model update based on lie algebra. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 728–735. IEEE, 2006.

[38] 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. *MFCA*, 2015.

[39] A. Srivastava and E. Klassen. Bayesian and geometric subspace tracking. *Advances in Applied Probability*, 36(01):43–56, 2004.

[40] J. Su, S. Kurtek, E. Klassen, A. Srivastava, et al. Statistical analysis of trajectories on riemannian manifolds: bird migration, hurricane tracking and video surveillance. *The Annals of Applied Statistics*, 8(1):530–552, 2014.

[41] H. Tabia, H. Laga, D. Picard, and P.-H. Gosselin. Covariance descriptors for 3d shape matching and retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4185–4192, 2014.

[42] B. S. Thomas. *Linear Subspace and Manifold Learning via Extrinsic Geometry*. PhD thesis, Duke University, 2015.

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

[44] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. In *European conference on computer vision*, pages 589–600. Springer, 2006.

[45] A. Tyagi and J. W. Davis. A recursive filter for linear systems on riemannian manifolds. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[46] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.

[47] A. Varentsova, S. Zhang, and K. Arfanakis. Development of a high angular resolution diffusion imaging human brain template. *NeuroImage*, 91:177–186, 2014.

[48] H. Zhang, P. A. Yushkevich, D. C. Alexander, et al. Deformable registration of diffusion tensor MR images with explicit orientation optimization. *MIA*, 10(5):764–785, 2006.