Nested Grassmanns for Dimensionality Reduction with Applications to Shape Analysis*

Chun-Hao Yang[1][0000−0002−2522−5957] and Baba C. Vemuri[2]**

[1] Department of Statistics, University of Florida, FL, USA
[2] Department of CISE, University of Florida, FL, USA

Abstract. Grassmann manifolds have been widely used to represent the geometry of feature spaces in a variety of problems in medical imaging and computer vision including but not limited to shape analysis, action recognition, subspace clustering and motion segmentation. For these problems, the features usually lie in a very high-dimensional Grassmann manifold and hence an appropriate dimensionality reduction technique is called for in order to curtail the computational burden. To this end, the Principal Geodesic Analysis (PGA), a nonlinear extension of the well known principal component analysis, is applicable as a general tool to many Riemannian manifolds. In this paper, we propose a novel framework for dimensionality reduction of data in Riemannian homogeneous spaces and then focus on the Grassman manifold which is an example of a homogeneous space. Our framework explicitly exploits the geometry of the homogeneous space yielding reduced dimensional nested sub-manifolds that need not be geodesic submanifolds and thus are more expressive. Specifically, we project points in a Grassmann manifold to an embedded lower dimensional Grassmann manifold. A salient feature of our method is that it leads to higher expressed variance compared to PGA which we demonstrate via synthetic and real data experiments.

Keywords: Grassmann manifolds · dimensionality reduction · shape analysis.

1 Introduction

In medical imaging, non-Euclidean spaces are commonly used to model descriptive features extracted from the raw data. For example, diffusion tensor imaging (DTI) [5] uses symmetric positive-definite (SPD) matrices to characterize the local diffusivity of the water molecules within each voxel of the DTI scan. The ensemble average propagator (EAP) [6], which captures the distribution of the diffusion of the water molecules in the tissue being imaged and is much more expressive than the diffusion tensors, can be parameterized as a point on a Hilbert sphere which is a constant curvature manifold. Another example is the Kendall's shape space (Procrustes shape space) [18] which is a manifold used to model shapes in computational anatomy. In the examples described above, the dataset is a collection of DTI scans, EAP fields, or the shapes of a segmented region of the brain. The data lie in a high-dimensional space: a DTI scan contains hundreds of thousands of diffusion tensors (DTs), which are 3×3 SPD matrices; an EAP

* This research was in part funded by the NSF grant IIS-1724174 to Vemuri. This paper will appear in the Proc. of IPMI 2021.
** Corresponding author, E-mail: vemuri@ufl.edu

field contains hundreds of thousands of EAPs and each EAP can be represented as a high-dimensional probability vector; the shape of the Corpus Callosum (which is used in our experiments) is represented by a several hundreds of boundary points in $\mathbb{R}^2$. Thus, in these cases, dimensionality reduction techniques, if applied appropriately, can benefit the subsequent statistical analysis.

Principal component analysis (PCA) is the simplest and most well-known (unsupervised) dimensionality reduction technique for data in $\mathbb{R}^n$. Using PCA, the data in $\mathbb{R}^n$ is projected to a vector subspace of dimension $k \ll n$ such that maximal variance in the original data is captured in the projected data. There are different generalizations of PCA to Riemannian manifolds. Authors in [11] proposed an expression for exact PGA, but due to the computational challenges they resorted to the tangent PCA (tPCA) approximation. Exact PGA (EPGA) was later proposed by [23] which does not use the tPCA approximation. However, EPGA is computationally expensive since it involves two non-linear optimizations steps per iteration (projection to the geodesic submanifold and finding the new geodesic direction such that the loss of information is minimized). Authors in [7] improved upon EPGA by deriving the closed-form expressions for the projection in the case of constant curvature manifolds, e.g. hyperspheres and hyperbolic spaces. There are several other variants of PGA, see [3,15,16,26]. Instead of projecting data to a geodesic submanifold, one may also find a curve on the manifold, called the principal curve [14] (this is a generalization of the principal curve in the Euclidean space by [13]), to represent the data using a lower dimensional submanifold.

PGA and its variants provided a dimensionality reduction technique for general Riemannian manifolds. Nonetheless, different Riemannian manifolds possess different geometric structures, e.g. curvature and symmetry. Therefore, by exploiting the geometry and/or other properties, one may design a more efficient and better dimensionality reduction method for specific Riemannian manifolds. For example, by utilizing the fact that S^q embedded inside S^p with $q < p$ is a geodesic submanifold of S^p, authors in [17] proposed a method called the *principal nested spheres* to perform dimensionality reduction on S^p. By translating the nested spheres, PGA on S^p can be seen as a special case of principal nested spheres. Another example is that of the manifold of SPD matrices, P_n. In [12], authors proposed to project data on P_n to P_m where $m \ll n$ by designing a projection map from P_n to P_m that maximized the projected variance or inter-class discrimination in the case of supervised dimensionality reduction. Although in this case, P_m is not a geodesic submanifold of P_n which makes it different from PGA, such an algorithm has the ability to handle supervised dimensionality reduction which PGA lacks. Indeed, considering only the geodesic submanifolds is certainly not sufficient in many applications. In [9], authors consider nested sequence of relations which lead to a nested sequence of submanifolds that are not necessarily geodesic. This works for most submanifolds of the Euclidean space, for example the n-sphere, as they are characterized by the set of solutions to some equations. However, for manifolds other than Euclidean spaces and spheres, the sequence of relations is not known and is probably nontrivial. Recently, Pennec proposed the exponential barycentric subspace (EBS) as a generalization of PGA [22]. A k-dimensional EBS is defined as the locus of weighted exponential barycenters of $(k + 1)$ affinely independent reference points. The nested structure of EBS can be achieved by adding or removing points. However,

the choice of the $(k+1)$ reference points can be computationally inefficient for high-dimensional spaces. Additional analysis is required to mitigate this problem.

Motivated by the expressiveness of the nested representations as observed in the earlier work described above, in this paper, we propose to develop a nested structure for Riemannian homogeneous spaces (under some mild conditions). Specifically, we will focus our attention on unsupervised and supervised dimensionality reduction for data on the Grassmann manifold $\mathrm{Gr}(p, V)$ which is a homogeneous space of all p-dimensional linear subspaces of the vector space V where $1 \leq p \leq \dim V$. We will assume that V is either $\mathbb{R}^n$ or $\mathbb{C}^n$. In shape analysis, the space of planar shapes, i.e., shapes that are represented by k ordered points in $\mathbb{R}^2$, is a complex projective space $\mathbb{C}P^{k-2} \cong \mathrm{Gr}(1, \mathbb{C}^{k-1})$. The number of points k is usually a few hundred and dimension of the underlying manifold $\mathrm{Gr}(1, \mathbb{C}^{k-1})$ is also large. Hence the core idea of our dimensionality reduction is to approximate $\mathcal{X} \in \mathrm{Gr}(p, V)$ by $\hat{\mathcal{X}} \in \mathrm{Gr}(p, \tilde{V})$ where $\dim \tilde{V} \ll \dim V$. Thus, the main contributions of our work here are: (i) We propose a nested structure for Riemannian homogeneous spaces (under some mild conditions) unifying the recently proposed nested structures for spheres [17] and SPD manifolds [12]. (ii) We present novel dimensionality reduction techniques based on the concept of nested geometric structures for the Grassmann manifold case. (iii) Synthetic and real data experiments demonstrate higher expressed variance captured by our lower dimensional nested submanifold representation scheme.

The rest of the paper is organized as follows. In Section 2, we review the geometry of the Grassmann manifold and present the nested structure of homogeneous spaces. Then, by using this nested structure, we describe algorithms for our unsupervised and supervised dimensionality reduction techniques for data on the Grassmann. In Section 3, we present experimental results. Finally, we draw conclusions in Section 4.

2 Nested Grassmannians

We will first review the Riemannian geometry of the Grassmann manifold in Section 2.1 and then the nested Grassmann model will be derived in Section 2.2. In Section 2.3 and 2.4, we describe the unsupervised and supervised dimensionality reduction using the nested Grassmann model. In Section 2.5, we will present some technical details required for implementation. A technique for the choosing the dimension of the 'reduced' model is then presented in Section 2.6.

2.1 The Riemannian Geometry of Grassmann Manifolds

For the sake of simplicity, we assume $V = \mathbb{R}^n$. For the case of $V = \mathbb{C}^n$, the results hold by replacing real matrices with complex matrices, $\boldsymbol{M}^T$ with the conjugate transpose $\boldsymbol{M}^H$, and the orthogonal group $\mathrm{O}(n)$ with the unitary group $\mathrm{U}(n)$. The *Grassmann manifold* $\mathrm{Gr}(p, n) := \mathrm{Gr}(p, \mathbb{R}^n)$ is the space of all p-dimensional subspaces in $\mathbb{R}^n$. The dimension of $\mathrm{Gr}(p, n)$ is $p(n-p)$. In this paper, for elements $\mathcal{X} \in \mathrm{Gr}(p, n)$, we write $\mathcal{X} = \mathrm{span}(\boldsymbol{X})$ where, $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p]$ is an orthonormal basis (o.n.b) for $\mathcal{X}$. The *Stiefel manifold* $\mathrm{St}(p, n)$ is the space of all orthonormal p-frames in $\mathbb{R}^n$. Let $\mathrm{O}(n) := \mathrm{St}(n, n)$ be the set of $n \times n$ orthogonal matrices. The Stiefel manifold can be identified as a *homogeneous space* $\mathrm{St}(p, n) \cong \mathrm{O}(n)/\mathrm{O}(n-p)$ and so is the Grassmann manifold $\mathrm{Gr}(p, n) \cong \mathrm{St}(p, n)/\mathrm{O}(p) \cong \mathrm{O}(n)/(\mathrm{O}(n) \times \mathrm{O}(n-p))$ [10]. There

are other ways to represent the Grassmann manifolds, see [19, Table 2] and we choose this particular one as it is the most widely used and easy to interpret. Since $\mathrm{St}(p,n)$ is a submanifold of $\mathbb{R}^{n\times p}$, a natural Riemannian metric for $\mathrm{St}(p,n)$ is induced from the Euclidean metric on $\mathbb{R}^{n\times p}$, i.e. for $\boldsymbol{U},\boldsymbol{V}\in T_{\boldsymbol{X}}\mathrm{St}(p,n)$, $\langle \boldsymbol{U},\boldsymbol{V}\rangle_{\boldsymbol{X}} = \mathrm{tr}(\boldsymbol{U}^T\boldsymbol{V})$. The canonical Riemannian metric on the Grassmann manifold is then inherited from the metric on $\mathrm{St}(p,n)$ as it is invariant to the left multiplication by elements of $O(n)$ [1,10]. We now state a few important geometric concepts that are relevant to our work.

With the canonical metric on the Grassmann manifolds, the geodesic can be expressed in closed form. Let $\mathcal{X} = \mathrm{span}(\boldsymbol{X}) \in \mathrm{Gr}(p,n)$ where $\boldsymbol{X}\in\mathrm{St}(p,n)$ and $\boldsymbol{H}$ be an $n\times p$ matrix. The the geodesic $\gamma(t)$ with $\gamma(0)=\mathcal{X}$ and $\gamma'(0)=\boldsymbol{H}$ is given by $\gamma_{\mathcal{X},\boldsymbol{H}}(t) = \mathrm{span}(\boldsymbol{X}\boldsymbol{V}\cos\boldsymbol{\Sigma}t + \boldsymbol{U}\sin\boldsymbol{\Sigma}t)$ where, $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ is the compact singular value decomposition of $\boldsymbol{H}$ [10, Theorem 2.3]. The *exponential map* at $\mathcal{X}$ is a map from $T_{\mathcal{X}}\mathrm{Gr}(p,n)$ to $\mathrm{Gr}(p,n)$ defined by $\mathrm{Exp}_{\mathcal{X}}\boldsymbol{H} = \gamma_{\mathcal{X},\boldsymbol{H}}(1) = \mathrm{span}(\boldsymbol{X}\boldsymbol{V}\cos\boldsymbol{\Sigma} + \boldsymbol{U}\sin\boldsymbol{\Sigma})$. If $\boldsymbol{X}^T\boldsymbol{Y}$ is invertible, the geodesic distance between $\mathcal{X}=\mathrm{span}(\boldsymbol{X})$ and $\mathcal{Y}=\mathrm{span}(\boldsymbol{Y})$ is given by $d_g^2(\mathcal{X},\mathcal{Y}) = \mathrm{tr}\,\boldsymbol{\Theta}^2 = \sum_{i=1}^p \theta_i^2$ where $(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^T)\boldsymbol{Y}(\boldsymbol{X}^T\boldsymbol{Y})^{-1} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, $\boldsymbol{U}\in\mathrm{St}(p,n)$, $\boldsymbol{V}\in\mathrm{O}(p)$, and $\boldsymbol{\Theta}=\tan^{-1}\boldsymbol{\Sigma}$. The diagonal entries $\theta_1,\ldots,\theta_k$ of $\boldsymbol{\Theta}$ are known as the principal angles.

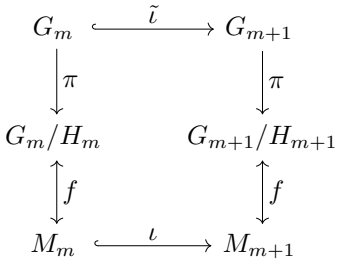

Fig. 1. Commutative diagram of the induced embedding for homogeneous spaces.

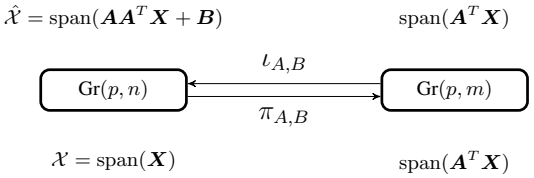

Fig. 2. Illustration of the embedding of $\mathrm{Gr}(p,m)$ in $\mathrm{Gr}(p,n)$ parameterized by $\boldsymbol{A}\in\mathrm{St}(m,n)$ and $\boldsymbol{B}\in\mathbb{R}^{n\times p}$ such that $\boldsymbol{A}^T\boldsymbol{B}=0$.

2.2 Embedding of $\mathrm{Gr}(p,m)$ in $\mathrm{Gr}(p,n)$

Let $\mathcal{X}=\mathrm{span}(\boldsymbol{X})\in\mathrm{Gr}(p,m)$, $\boldsymbol{X}\in\mathrm{St}(p,m)$. The map $\iota:\mathrm{Gr}(p,m)\to\mathrm{Gr}(p,n)$, for $m<n$, defined by, $\iota(\mathcal{X}) = \mathrm{span}\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{0}_{(n-m)\times p}\end{bmatrix}\right)$ is an embedding and it is easy to check that this embedding is isometric [25, Eq. (8)]. However, for the dimensionality reduction problem, the above embedding is insufficient as it is not flexible enough to encompass other possible embeddings. To design flexible embeddings, we propose a general framework for Riemannian homogeneous spaces $M\cong G/H$ e.g., the Grassmann manifold, see Figure 1. Note that in Figure 1, we impose the following **mild conditions:** *(i) the groups G_m and G_{m+1} are of the same "type", i.e., $G_m = GL(m)$ or $G_m = O(m)$, and so are H_m and H_{m+1}, and (ii) the diagram commutes.*

The idea is to define an embedding $\tilde{\iota}:G_m\to G_{m+1}$. This embedding $\tilde{\iota}$, together with the submersion $\pi:G_m\to G_m/H_m$ and the identification map $f:G_m/H_m\to M_m$, induces an embedding $\iota:M_m\to M_{m+1}$ provided that *the diagram commutes.*

For the Grassmann manifolds, $G_m = \mathrm{O}(m)$ and $H_m = \mathrm{O}(m) \times \mathrm{O}(m-p)$. We consider the embedding $\tilde{\iota}_m : \mathrm{O}(m) \to \mathrm{O}(m+1)$ given by,

$$\tilde{\iota}_m(\boldsymbol{O}) = \mathrm{GS}\left(\boldsymbol{R}\begin{bmatrix}\boldsymbol{O} & \boldsymbol{a}\\ \boldsymbol{b}^T & c\end{bmatrix}\right) \tag{1}$$

where $\boldsymbol{O} \in \mathrm{O}(m)$, $\boldsymbol{R} \in \mathrm{O}(m+1)$, $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m$, $c \in \mathbb{R}$, $c \neq \boldsymbol{b}^T\boldsymbol{O}^{-1}\boldsymbol{a}$, and $\mathrm{GS}(\cdot)$ is the Gram-Schmidt process. Hence the induced embedding $\iota_m : \mathrm{Gr}(p,m) \to \mathrm{Gr}(p,m+1)$ is given by,

$$\iota_m(\mathcal{X}) = \mathrm{span}\left(\boldsymbol{R}\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{b}^T\end{bmatrix}\right) = \mathrm{span}(\tilde{\boldsymbol{R}}\boldsymbol{X} + \boldsymbol{v}\boldsymbol{b}^T),$$

where $\boldsymbol{b} \in \mathbb{R}^p$, $\boldsymbol{R} \in \mathrm{O}(m+1)$, $\tilde{\boldsymbol{R}}$ contains the first m columns of $\boldsymbol{R}$ (which means $\tilde{\boldsymbol{R}} \in \mathrm{St}(m, m+1)$), $\boldsymbol{v}$ is the last column of $\boldsymbol{R}$, and $\mathcal{X} = \mathrm{span}(\boldsymbol{X}) \in \mathrm{Gr}(p,m)$. It is easy to see that for $\boldsymbol{R} = \boldsymbol{I}$ and $\boldsymbol{b} = \boldsymbol{0}$, this gives the natural embedding described at the beginning of this section.

Proposition 1 *If $\boldsymbol{b} = 0$, then ι_m is an isometric embedding.*

Proof. Let $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{R}}\boldsymbol{X}$ and $\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{R}}\boldsymbol{Y}$. We first compute the principal angles between $\mathrm{span}(\tilde{\boldsymbol{X}})$ and $\mathrm{span}(\tilde{\boldsymbol{Y}})$. Since $\tilde{\boldsymbol{R}} \in \mathrm{St}(m, m+1)$, we have $(\boldsymbol{I} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T)\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{Y}})^{-1} = \tilde{\boldsymbol{R}}(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^T)\boldsymbol{Y}(\boldsymbol{X}^T\boldsymbol{Y})^{-1}$. Hence the principal angles between $\mathrm{span}(\tilde{\boldsymbol{X}})$ and $\mathrm{span}(\tilde{\boldsymbol{Y}})$ are the same as those of $\mathrm{span}(\boldsymbol{X})$ and $\mathrm{span}(\boldsymbol{Y})$. By the Myers-Steenrod theorem [21, Theorem 1], ι_m is an isometric embedding.

With the embedding ι_m, we can construct the corresponding projection $\pi_m : \mathrm{Gr}(p, m+1) \to \mathrm{Gr}(p,m)$ using the following proposition.

Proposition 2 *The projection $\pi_m : Gr(p, m+1) \to Gr(p,m)$ corresponding to ι_m is given by $\pi_m(\mathcal{X}) = \mathrm{span}(\tilde{\boldsymbol{R}}^T\boldsymbol{X})$.*

Proof. First, let $\mathcal{Y} = \mathrm{span}(\boldsymbol{Y}) \in \mathrm{Gr}(p,m)$ and $\mathcal{X} = \mathrm{span}(\boldsymbol{X}) \in \mathrm{Gr}(p,m+1)$ be such that $\mathcal{X} = \mathrm{span}(\tilde{\boldsymbol{R}}\boldsymbol{Y} + \boldsymbol{v}\boldsymbol{b}^T)$. Then $\boldsymbol{X}\boldsymbol{L} = \tilde{\boldsymbol{R}}\boldsymbol{Y} + \boldsymbol{v}\boldsymbol{b}^T$ for some $\boldsymbol{L} \in \mathrm{GL}(p)$. Therefore, $\boldsymbol{Y} = \tilde{\boldsymbol{R}}^T(\boldsymbol{X}\boldsymbol{L} - \boldsymbol{v}\boldsymbol{b}^T) = \tilde{\boldsymbol{R}}^T\boldsymbol{X}\boldsymbol{L}$ and $\mathcal{Y} = \mathrm{span}(\boldsymbol{Y}) = \mathrm{span}(\tilde{\boldsymbol{R}}^T\boldsymbol{X}\boldsymbol{L}) = \mathrm{span}(\tilde{\boldsymbol{R}}^T\boldsymbol{X})$. Hence, the projection is given by $\pi_m(\mathcal{X}) = \mathrm{span}(\tilde{\boldsymbol{R}}^T\boldsymbol{X})$. This completes the proof. ∎

The nested relation can be extended inductively and we refer to this construction as the *nested Grassmann structure*:

$$\mathrm{Gr}(p,m) \overset{\iota_m}{\hookrightarrow} \mathrm{Gr}(p,m+1) \overset{\iota_{m+1}}{\hookrightarrow} \ldots \overset{\iota_{n-2}}{\hookrightarrow} \mathrm{Gr}(p,n-1) \overset{\iota_{n-1}}{\hookrightarrow} \mathrm{Gr}(p,n).$$

Thus the embedding from $\mathrm{Gr}(p,m)$ into $\mathrm{Gr}(p,n)$ can be constructed inductively by $\iota \coloneqq \iota_{n-1} \circ \ldots \circ \iota_{m-1} \circ \iota_m$ and similarly for the corresponding projection. The explicit forms of the embedding and the projection are given in the following proposition.

Proposition 3 *The embedding of $Gr(p,m)$ into $Gr(p,n)$ for $m < n$ is given by $\iota_{\boldsymbol{A},\boldsymbol{B}}(\mathcal{X}) = \mathrm{span}(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{B})$ where $\boldsymbol{A} \in St(m,n)$ and $\boldsymbol{B} \in \mathbb{R}^{n\times p}$ such that $\boldsymbol{A}^T\boldsymbol{B} = 0$. The corresponding projection from $Gr(p,n)$ to $Gr(p,m)$ is given by $\pi_{\boldsymbol{A}} = \mathrm{span}(\boldsymbol{A}^T\boldsymbol{X})$.*

Proof. By the definition, $\iota := \iota_{n-1} \circ \ldots \circ \iota_{m-1} \circ \iota_m$ and thus the embedding $\iota : \mathrm{Gr}(p,m) \to \mathrm{Gr}(p,n)$ can be simplified as

$$\iota_{A,B}(\mathcal{X}) = \mathrm{span}\left(\left(\prod_{i=m}^{n-1} \boldsymbol{R}_i\right)\boldsymbol{X} + \sum_{i=m}^{n-1}\left(\prod_{j=i+1}^{n-1} \boldsymbol{R}_j\right)\boldsymbol{v}_i\boldsymbol{b}_i^T\right) = \mathrm{span}(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{B})$$

where $\boldsymbol{R}_i \in \mathrm{St}(i, i+1)$, $\boldsymbol{v}_i$ is such that $[\boldsymbol{R}_i \ \ \boldsymbol{v}_i] \in \mathrm{O}(i+1)$, $\boldsymbol{b}_i \in \mathbb{R}^p$, $\boldsymbol{A} = \boldsymbol{R}_{n-1}\boldsymbol{R}_{n-2}\ldots\boldsymbol{R}_m \in \mathrm{St}(m,n)$, and $\boldsymbol{B} = \sum_{i=m}^{n-1}\left(\prod_{j=i+1}^{n-1}\boldsymbol{R}_j\right)\boldsymbol{v}_i\boldsymbol{b}_i^T$ is an $n \times p$ matrix. It is easy to see that $\boldsymbol{A}^T\boldsymbol{B} = 0$. Similar to Proposition 2, the projection $\pi_{\boldsymbol{A}} : \mathrm{Gr}(p,n) \to \mathrm{Gr}(p,m)$ is then given by $\pi_{\boldsymbol{A}}(\mathcal{X}) = \mathrm{span}(\boldsymbol{A}^T\boldsymbol{X})$. This completes the proof. ■

From Proposition 1, if $\boldsymbol{B} = 0$ then $\iota_{\boldsymbol{A}}$ is an isometric embedding. *Hence, our nested Grassmann structure is more flexible than PGA as it allows one to project the data onto a non-geodesic submanifold.* An illustration is shown in Figure 2.

Connections to Other Nested Structures The nested homogeneous spaces proposed in this work (see Figure 1) is actually a unifying framework for the nested spheres [17] and the nested SPD manifolds [12]. Since the n-sphere can be identified with a homogeneous space $S^{n-1} \cong \mathrm{O}(n)/\mathrm{O}(n-1)$, with the embedding (1), the induced embedding of S^{n-1} into S^n is

$$\iota(\boldsymbol{x}) = \mathrm{GS}\left(\boldsymbol{R}\begin{bmatrix}\boldsymbol{x}\\ b\end{bmatrix}\right) = \frac{1}{\sqrt{1+b^2}}\boldsymbol{R}\begin{bmatrix}\boldsymbol{x}\\ b\end{bmatrix} = \boldsymbol{R}\begin{bmatrix}\sin(r)\boldsymbol{x}\\ \cos(r)\end{bmatrix}$$

where $\boldsymbol{x} \in S^{n-1}$, $b \in \mathbb{R}$, and $r = \cos^{-1}\left(\frac{b}{\sqrt{1+b^2}}\right)$. This is precisely the nested sphere proposed in [17, Eq. (2)]. For the SPD manifold $P_n \cong \mathrm{GL}(n)/\mathrm{O}(n)$, a carefully chosen embedding of $\mathrm{GL}(n)$ into $\mathrm{GL}(n+1)$ gives the nested SPD manifolds in [12].

2.3 Unsupervised Dimensionality Reduction

We can now apply the nested Grassmann (NG) structure to the problem of unsupervised dimensionality reduction. Suppose that we are given the points, $\mathcal{X}_1, \ldots, \mathcal{X}_N \in \mathrm{Gr}(p,n)$. We would like to have lower dimensional representations in $\mathrm{Gr}(p,m)$ for $\mathcal{X}_1, \ldots, \mathcal{X}_N$ with $m \ll n$. The desired projection map $\pi_{\boldsymbol{A}}$ that we seek is obtained by the minimizing the reconstruction error, i.e. $L_u(\boldsymbol{A},\boldsymbol{B}) = \frac{1}{N}\sum_{i=1}^N d^2(\mathcal{X}_i, \hat{\mathcal{X}}_i)$ where d is a distance metric on $\mathrm{Gr}(p,n)$. It is clear that L_u has a $\mathrm{O}(m)$-symmetry in the first argument, i.e. $L_u(\boldsymbol{A}\boldsymbol{O},\boldsymbol{B}) = L_u(\boldsymbol{A},\boldsymbol{B})$ for $\boldsymbol{O} \in \mathrm{O}(m)$. Hence, the optimization is performed over the space $\mathrm{St}(m,n)/\mathrm{O}(m) \cong \mathrm{Gr}(m,n)$ when optimizing with respect to this particular loss function. Now we can apply the Riemannian gradient descent algorithm [10] to obtain $\boldsymbol{A}$ and $\boldsymbol{B}$ by optimizing $L_u(\boldsymbol{A},\boldsymbol{B})$ over $\mathrm{span}(\boldsymbol{A}) \in \mathrm{Gr}(m,n)$ and $\boldsymbol{B} \in \mathbb{R}^{n\times p}$ such that $\boldsymbol{A}^T\boldsymbol{B} = 0$. Note that the restriction $\boldsymbol{A}^T\boldsymbol{B} = 0$ simply means that the columns of $\boldsymbol{B}$ are in the nullspace of $\boldsymbol{A}^T$, denoted $N(\boldsymbol{A}^T)$. Hence in practice this restriction can be handled as follows: For arbitrary $\tilde{\boldsymbol{B}} \in \mathbb{R}^{n\times p}$, project $\tilde{\boldsymbol{B}}$ on to $N(\boldsymbol{A}^T)$, i.e. $\boldsymbol{B} = P_{N(\boldsymbol{A}^T)}\tilde{\boldsymbol{B}}$ where $P_{N(\boldsymbol{A}^T)} = \boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^T$ is the projection from $\mathbb{R}^n$ to $N(\boldsymbol{A}^T)$. Thus, the loss function can be written as

$$L_u(\boldsymbol{A},\boldsymbol{B}) = \frac{1}{N}\sum_{i=1}^N d^2(\mathrm{span}(\boldsymbol{X}_i), \mathrm{span}(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{X}_i + (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^T)\boldsymbol{B}))$$

and it is optimized over $\mathrm{Gr}(m,n) \times \mathbb{R}^{n\times p}$.

2.4 Supervised Dimensionality Reduction

If in addition to $\mathcal{X}_1,\ldots,\mathcal{X}_N \in \mathrm{Gr}(p,n)$, we are given the associated labels $y_1,\ldots,y_N \in \{1,\ldots,k\}$, then we would like to utilize this extra information to sharpen the result of dimensionality reduction. Specifically, we expect that after reducing the dimension, points from the same class are still close to each other while points from different classes are well separated. We use an *affinity function* $a : \mathrm{Gr}(p,n) \times \mathrm{Gr}(p,n) \to \mathbb{R}$ to encode the structure of the data as suggested by [12, Sec 3.1, Eq. (14)-(16)]. The desired projection map π_A that we seek is obtained by the minimizing the following loss function

$$L_s(\boldsymbol{A}) = \frac{1}{N^2}\sum_{i,j=1}^{N} a(\mathcal{X}_i,\mathcal{X}_j)d^2(\mathrm{span}(\boldsymbol{A}^T\boldsymbol{X}_i),\mathrm{span}(\boldsymbol{A}^T\boldsymbol{X}_j))$$

where d is a distance metric on $\mathrm{Gr}(p,m)$. Note that if the distance metric d has $\mathrm{O}(m)$-symmetry, e.g. the geodesic distance, so does L_s. In this case the optimization can be done on $\mathrm{St}(m,n)/\mathrm{O}(m) \cong \mathrm{Gr}(m,n)$. Otherwise it is on $\mathrm{St}(m,n)$. This supervised dimensionality reduction is termed as, supervised nested Grassmann (sNG).

2.5 Choice of the distance d

The loss functions L_u and L_s depend on the choice of the distance $d : \mathrm{Gr}(p,n) \times \mathrm{Gr}(p,n) \to \mathbb{R}_{\geq 0}$. Besides the geodesic distance, there are many widely used distances on the Grassmann manifold, see for example [10, p. 337] and [25, Table 2]. In this work, we use two different distance metrics: (1) the geodesic distance d_g and (2) the projection distance, which is also called the chordal distance in [25] and the projection F-norm in [10]. The geodesic distance was defined in Section 2.1 and the projection distance is defined as follows. For $\mathcal{X},\mathcal{Y} \in \mathrm{Gr}(p,n)$, denote the projection matrices onto $\mathcal{X}$ and $\mathcal{Y}$ by $P_\mathcal{X}$ and $P_\mathcal{Y}$ respectively. Then, the distance between $\mathcal{X}$ and $\mathcal{Y}$ is given by $d_p(\mathcal{X},\mathcal{Y}) = \|P_\mathcal{X} - P_\mathcal{Y}\|_F/\sqrt{2} = \left(\sum_{i=1}^{p}\sin^2\theta_i\right)^{1/2}$ where $\theta_1,\ldots,\theta_p$ are the principal angles of $\mathcal{X}$ and $\mathcal{Y}$. If $\mathcal{X} = \mathrm{span}(X)$, then $P_\mathcal{X} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$. It is also easy to see the the projection distance has $\mathrm{O}(n)$-symmetry. We choose the projection distance mainly for its computational efficiency as it involves only matrix multiplication which has a time complexity $O(n^2)$ while the geodesic distance requires SVD which has a time complexity of $O(n^3)$.

2.6 Analysis of Principal Nested Grassmanns

In practice, we might not have prior knowledge about m. So one can choose $p < m_1 < \ldots < m_k < n$ and construct a sequence of Grassmann manifolds. Then, for each nested Grassmann, we compute the percentage of variance explained. Suppose $\mathcal{X}_1 = \mathrm{span}(\boldsymbol{X}_1),\ldots,\mathcal{X}_N = \mathrm{span}(\boldsymbol{X}_N) \in \mathrm{Gr}(p,n)$ and $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are obtained for $\mathrm{Gr}(p,m_i)$ from the algorithm described in the previous section. The percentage of variance explained in $\mathrm{Gr}(p,m_i)$ is given by the ratio of variance of the $\hat{\mathcal{X}}_j$'s where $\hat{\mathcal{X}}_j = \mathrm{span}(\boldsymbol{A}_i^T\boldsymbol{X}_j)$ and the variance of the $\mathcal{X}_j$'s. The desired dimension m can be chosen according to the desired percentage of variance explained somewhat similar to the way one chooses the number of principal components.

3 Experiments

In this section, we will demonstrate the performance of the proposed dimensionality reduction technique, i.e. NG and sNG, via experiments on synthetic and real data. The implementation[3] is based on the python library `pymanopt` [24] and we use the conjugate gradient descent algorithm for the optimization (with default parameters in pymanopt). The optimization was performed on a desktop with 3.6GHz Intel i7 processors and took about 30 seconds to converge.

3.1 Synthetic Data

In this subsection, we compare the performance of the projection and the geodesic distances respectively. The questions we will answer are the following: (1) From Section 2.5, we see that using projection distance is more efficient than using the geodesic distance. But how do they perform compared to each other under varying dimension n and variance level σ^2? (2) Is our method of dimensionality reduction better than PGA? Under what conditions does our method outperform PGA?

Comparison of Projection and Geodesic Distances The procedure we used to generate random points on $\mathrm{Gr}(p,n)$ for the synthetic experiments is the following. First, we generate N points from a uniform distribution on $\mathrm{St}(p,m)$ [8, Ch. 2.5], generate A from the uniform distribution on $\mathrm{St}(m,n)$, and generate B as an $n \times p$ matrix with i.i.d entries from $N(0, 0.1)$. Then we compute $\tilde{\mathcal{X}}_i = \mathrm{span}(\boldsymbol{A}\boldsymbol{X}_i + (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^T)\boldsymbol{B}) \in \mathrm{Gr}(p,n)$. Finally, we compute $\mathcal{X}_i = \mathrm{Exp}_{\tilde{\mathcal{X}}_i}(\sigma \boldsymbol{U}_i)$, where $\boldsymbol{U}_i = \tilde{\boldsymbol{U}}_i / \|\tilde{\boldsymbol{U}}_i\|$ and $\tilde{\boldsymbol{U}}_i \in T_{\tilde{\mathcal{X}}_i}\mathrm{Gr}(p,n)$, to include some perturbation.

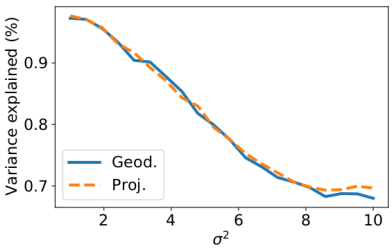

Fig. 3. Comparison of the NG representations based on the projection and geodesic distances using the ratio of expressed variance.

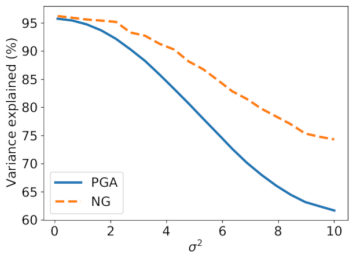

Fig. 4. Comparison of NG and PGA algorithms via percentage of explained variance.

This experiment involves comparing the performance of the NG representation in terms of the ratio of the variance explained, under different levels of data variance. In this experiment, we set $N = 50$, $n = 10$, $m = 3$, and $p = 1$ and σ is ranging from 1 to 10. The results are averaged over 100 repetitions and are shown in Figure 3. From these results, we can see that the ratios of variance explained for the projection distance and the geodesic distance are indistinguishable but the one using projection distance is

[3] The code is available at https://github.com/cvgmi/NestedGrassmann.

much faster than the one using the geodesic distance. The reason is that when two points on the Grassmann manifold are close, the geodesic distance can be well-approximated by the projection distance. When the algorithm converges, the original point $\mathcal{X}_i$ and the reconstructed point $\hat{\mathcal{X}}_i$ should be close and the geodesic distance can thus be well-approximated by the projection distance. Therefore, for the experiments in the next section, we use the projection distance for the sake of efficiency.

Comparison of NG and PGA Now we compare our NG representation to PGA. Similar to the previous experiment, we set $N = 50$, $n = 30$, $m = 20$, $p = 2$, and $\sigma = 0.1$ and apply the same procedure to generate synthetic data. There is a subtle difference between PGA and NG, that is, in order to project the points on $\text{Gr}(p, n) = \text{Gr}(2, 30)$ to a $\tilde{m}$-dimensional submanifold, for PGA we need to choose $\tilde{m}$ principal components and for NG we need to project them to $\text{Gr}(2, \tilde{m}/2 + 2)$ (since $\dim \text{Gr}(2, \tilde{m}/2 + 2) = \tilde{m}$). The results are averaged over 100 repetitions and are shown in Table 1.

	$\tilde{m}$ (dim. of submanifold)				
	2	4	6	8	10
NG	33.12%	50.49%	59.98%	67.85%	73.77%
PGA	16.36%	29.41%	40.81%	50.63%	59.29%

Table 1. The percentage of explained variance by PGA and NG representations respectively.

Fig. 5. Example Corpus Callosi shapes from three distinct age groups, each depicted using the boundary point sets.

From Table 1, we can see that our method outperforms PGA by virtue of the fact that it is able to capture a larger amount of variance contained in the data. Next, we will investigate the conditions under which our method and PGA perform equally well and when our method outperforms PGA. To answer this question, we set $N = 50$, $n = 10$, $m = 5$, $p = 2$, and σ is ranging from 0.01 to 2. We then apply PGA and NG to reduce the dimension to 2 (i.e. choosing 2 principal components in PGA and project to $\text{Gr}(2, 3)$ in NG). The results are averaged over 100 repetitions and are shown in Figure 4. We can see that when the variance is small, our method produces almost the same result as PGA, whereas, our method is significantly better for the large data variance case. Note that when the variance in the data is small, i.e. the data are tightly clustered around the FM and PGA captures the essence of the data well. However, the requirement in PGA on the geodesic submanifold to pass through the anchor point, namely the FM, is not meaningful for data with large variance as explained through the following simple example. Consider, a few data points spread out on the equator of a sphere. The FM in this case is likely to be the north pole of the sphere if we restrict ourselves to the upper hemisphere. Thus, the geodesic submanifold computed by PGA will pass through this FM. However, what is more meaningful is a submanifold corresponding to the equator, which is what a nested spheres representation [17] in this case yields. In similar vein, for data with large variance on a Grassmann manifold, our NG representation will yield a more meaningful representation than PGA.

3.2 Application to Planar Shape Analysis

We now apply our method to planar (2-dimensional) shape analysis. A planar shape σ can be represented as an ordered set of $k > 2$ points in $\mathbb{R}^2$, called k-ads. Here we assume

that these k points are not all identical. The space of all planar shapes (after removing the effect of translations, rotations, and scaling) is denoted by $\Sigma_2^k = (\mathbb{R}^{k\times 2}/\text{Sim}(2)) \setminus \{0\}$ where $\text{Sim}(2)$ is the group of similarity transformations of $\mathbb{R}^2$, i.e. if $g \in \text{Sim}(2)$, then $g \cdot \boldsymbol{x} = s\boldsymbol{R}\boldsymbol{x} + \boldsymbol{t}$ for some $s > 0$, $\boldsymbol{R} \in \text{O}(2)$, and $\boldsymbol{t} \in \mathbb{R}^2$. $\{0\}$ is excluded because we assume the k points are not all identical. It was shown in [18] that Σ_2^k is a smooth manifold and, when equipped with the Procrustean metric, is isometric to the complex projective space $\mathbb{C}P^{k-2}$ equipped with the Fubini-Study metric which is a special case of the complex Grassmannians, i.e. $\mathbb{C}P^{k-2} \cong \text{Gr}(1, \mathbb{C}^{k-1})$. In practice, we need to preprocess the k-ads as follows to make it lie in $\text{Gr}(1, \mathbb{C}^{k-1})$. Let $\boldsymbol{X}$ be the $k \times 2$ matrix containing the k points in $\mathbb{R}^2$. First, the effect of translation is removed by subtracting the first point. Then all these points are mapped to the complex vector space and take the span of the resulting vector to remove the effect of rotation and scaling.

OASIS Corpus Callosum Data Experiment The OASIS database [20] is a publicly available database that contains T1-MR brain scans of subjects of age ranging from 18 to 96. In particular, it includes subjects that are clinically diagnosed with mild to moderate Alzheimers disease. We further classify them into three groups: *young* (aged between 10 and 40), *middle-aged* (aged between 40 and 70), and *old* (aged above 70). For demonstration, we randomly choose 4 brain scans within each decade, totalling 36 brain scans. From each scan, the Corpus Callosum (CC) region is segmented and 250 points are taken on the boundary of the CC region. See Figure 5 for samples of the segmented corpus callosi. In this case, the shape space is $\Sigma_2^{248} \cong \mathbb{C}P^{248} \cong \text{Gr}(1, \mathbb{C}^{249})$. Results are shown in Table 2. Note that in Table 2, m is the dimension of the submanifold, i.e. for NG, we project to $\text{Gr}(1, \mathbb{C}^{m+1})$ and for PGA, we take first m principal components.

	m				
	1	5	10	15	20
NG	26.38%	68.56%	84.18%	90.63%	94.04%
PGA	7.33%	43.74%	73.48%	76.63%	79.9%

Table 2. Percentage of explained variance by PGA and NG representations respectively.

	Accuracy	Explained Var.
gKNN	33.33%	N/A
gKNN + sPGA	38.89%	3.27%
gKNN + sNG	66.67%	98.7%
gKNN + PGA	30.56%	46.61%
gKNN + NG	30.56%	84.28%

Table 3. Classification accuracies and explained variances for sPGA and sNG.

Since the data are divided into three groups (young, middle-aged, and old), we can apply the sNG described in Section 2.4 to reduce the dimension. *The purpose of this experiment is not to demonstrate state-of-the-art classification accuracy for this dataset. Instead, our goal here is to demonstrate that the proposed nested Grassmann representation in a supervised setting is much more discriminative than the competition, namely the supervised PGA*. Hence, we choose a naive and impoverished classifier such as the geodesic kNN (gKNN) to highlight the aforementioned discriminative power of the nested Grassmann over PGA.

For comparison, the PGA can be easily extended to *supervised PGA* (sPGA) by first diffeomorphically mapping all the data to the tangent space anchored at the FM and then performing supervised PCA [2,4] on the tangent space. In this demonstration, we apply a gKNN classifier with $k = 5$ to the data before and after reducing the dimension (with and without supervision). Specifically, *the classification here is using*

a leave-one-out technique, i.e. the prediction of $\mathcal{X}_j$ is determined by the geodesic k nearest neighbors of the $\mathcal{X}_i$'s excluding $\mathcal{X}_j$. In this experiment, we choose $m = 10$, i.e. $\mathrm{Gr}(1, \mathbb{C}^{249}) \to \mathrm{Gr}(1, \mathbb{C}^{11})$ (for PGA/sPGA, the number of principal components would be $m = 10$). The results are shown in Table 3. These results are in accordance with our expectation since in both sNG and sPGA, we seek a projection that minimizes the within-group variance while maximizing the between-group variance. However, as we observed earlier, the constraint of requiring the geodesic submanifold to pass through the FM is not well suited for this dataset which has a large variance across the data. This accounts for why the sNG exhibits far superior performance compared to sPGA in accuracy as well as in explained variance.

4 Conclusion

In this work, we proposed a nested structure for homogeneous spaces and applied this structure to the dimensionality reduction problems of data in Grassmann manifolds. We also discuss how this nested structure served as a generalization of other existing nested structures for spheres and the manifold of SPD matrices. Specifically, we showed that a lower dimensional Grassmann manifold can be embedded into a higher dimensional Grassmann manifold and via this embedding we constructed a sequence of nested Grassmann manifolds. Compared to the PGA, which is designed for general Riemannian manifolds, the proposed method can capture a higher percentage of data variance after reducing the dimensionality. This is primarily because our method unlike the PGA does not require the reduced dimensional submanifold to be a geodesic submanifold. To sum up, the nested Grassmann structure allows us to fit the data to a larger class of submanifolds than PGA. We also proposed a supervised dimensionality reduction technique which simultaneously differentiates data classes while reducing dimensionality. Efficacy of our method was demonstrated on the OASIS Corpus Callosi data for dimensionality reduction and classification. We showed that our method outperforms the widely used PGA significantly.

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. Acta Applicandae Mathematica **80**(2), 199–220 (2004) 4
2. Bair, E., Hastie, T., Paul, D., Tibshirani, R.: Prediction by supervised principal components. Journal of the American Statistical Association **101**(473), 119–137 (2006) 10
3. Banerjee, M., Chakraborty, R., Vemuri, B.C.: Sparse exact PGA on Riemannian manifolds. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 5010–5018 (2017) 2
4. Barshan, E., Ghodsi, A., Azimifar, Z., Jahromi, M.Z.: Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. Pattern Recognition **44**(7), 1357–1371 (2011) 10
5. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophysical journal **66**(1), 259–267 (1994) 1
6. Callaghan, P.T.: Principles of Nuclear Magnetic Resonance Microscopy. Oxford University Press on Demand (1993) 1

7. Chakraborty, R., Seo, D., Vemuri, B.C.: An efficient exact-PGA algorithm for constant curvature manifolds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3976–3984 (2016) 2
8. Chikuse, Y.: Statistics on Special Manifolds, vol. 174. Springer Science & Business Media (2003) 8
9. Damon, J., Marron, J.: Backwards principal component analysis and principal nested relations. Journal of Mathematical Imaging and Vision **50**(1-2), 107–114 (2014) 2
10. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM Journal on Matrix Analysis and Applications **20**(2), 303–353 (1998) 3, 4, 6, 7
11. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging **23**(8), 995–1005 (2004) 2
12. Harandi, M., Salzmann, M., Hartley, R.: Dimensionality reduction on SPD manifolds: The emergence of geometry-aware methods. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(1), 48–62 (2018) 2, 3, 6, 7
13. Hastie, T., Stuetzle, W.: Principal curves. Journal of the American Statistical Association **84**(406), 502–516 (1989) 2
14. Hauberg, S.: Principal curves on Riemannian manifolds. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(9), 1915–1921 (2016) 2
15. Huckemann, S., Hotz, T., Munk, A.: Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. Statistica Sinica pp. 1–58 (2010) 2
16. Huckemann, S., Ziezold, H.: Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces. Advances in Applied Probability **38**(2), 299–319 (2006) 2
17. Jung, S., Dryden, I.L., Marron, J.: Analysis of principal nested spheres. Biometrika **99**(3), 551–568 (2012) 2, 3, 6, 9
18. Kendall, D.G.: Shape manifolds, Procrustean metrics, and complex projective spaces. Bulletin of the London Mathematical Society **16**(2), 81–121 (1984) 1, 10
19. Lai, Z., Lim, L.H., Ye, K.: Simpler Grassmannian optimization. arXiv preprint arXiv:2009.13502 (2020) 4
20. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience **19**(9), 1498–1507 (2007) 10
21. Myers, S.B., Steenrod, N.E.: The group of isometries of a Riemannian manifold. Annals of Mathematics pp. 400–416 (1939) 5
22. Pennec, X., et al.: Barycentric subspace analysis on manifolds. The Annals of Statistics **46**(6A), 2711–2746 (2018) 2
23. Sommer, S., Lauze, F., Hauberg, S., Nielsen, M.: Manifold valued statistics, exact principal geodesic analysis and the effect of linear approximations. In: European Conference on Computer Vision. pp. 43–56. Springer (2010) 2
24. Townsend, J., Koep, N., Weichwald, S.: Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation. Journal of Machine Learning Research **17**(137), 1–5 (2016), http://jmlr.org/papers/v17/16-177.html 8
25. Ye, K., Lim, L.H.: Schubert varieties and distances between subspaces of different dimensions. SIAM Journal on Matrix Analysis and Applications **37**(3), 1176–1197 (2016) 4, 7
26. Zhang, M., Fletcher, T.: Probabilistic principal geodesic analysis. In: Advances in Neural Information Processing Systems. pp. 1178–1186 (2013) 2