Nested Grassmannians for Dimensionality Reduction with Applications

Chun-Hao Yang   chunhaoy@ntu.edu.tw
Institute of Applied Mathematical Science, National Taiwan University, Taipei, Taiwan

Baba C. Vemuri   vemuri@ufl.edu
Department of CISE, University of Florida, Gainesville, FL, USA

Abstract

In the recent past, nested structures in Riemannian manifolds has been studied in the context of dimensionality reduction as an alternative to the popular principal geodesic analysis (PGA) technique, for example, the principal nested spheres. In this paper, we propose a novel framework for constructing a nested sequence of homogeneous Riemannian manifolds. Common examples of homogeneous Riemannian manifolds include the n-sphere, the Stiefel manifold, the Grassmann manifold and many others. In particular, we focus on applying the proposed framework to the Grassmann manifold, giving rise to the nested Grassmannians (NG). An important application in which Grassmann manifolds are encountered is planar shape analysis. Specifically, each planar (2D) shape can be represented as a point in the complex projective space which is a complex Grassmann manifold. Some salient features of our framework are: (i) it explicitly exploits the geometry of the homogeneous Riemannian manifolds and (ii) the nested lower-dimensional submanifolds need not be geodesic. With the proposed NG structure, we develop algorithms for the supervised and unsupervised dimensionality reduction problems respectively. The proposed algorithms are compared with PGA via simulation studies and real data experiments and are shown to achieve a higher ratio of expressed variance compared to PGA.

Keywords: Grassmann Manifolds, Dimensionality Reduction, Shape Analysis, Homogeneous Riemannian Manifolds

1. Introduction

Riemannian manifolds are often used to model the sample space in which features derived from the raw data encountered in many medical imaging applications live. Common examples include the diffusion tensors (DTs) in diffusion tensor imaging (DTI) (Basser et al., 1994), the ensemble average propagator (EAP) (Callaghan, 1993). Both DTs and EAP are used to capture the diffusional properties of water molecules in the central nervous system by non-invasively imaging the tissue via diffusion weighted magnetic resonance imaging. In DTI, diffusion at a voxel is captured by a DT, which is a 3×3 symmetric positive-definite matrix, whereas EAP is a probability distribution characterizing the local diffusion at a voxel, which can be parametrized as a point on the Hilbert sphere. Another example is the shape space used to represent shapes in shape analysis. There are many ways to represent a shape, and the most simple one is to use landmarks. For the landmark-based representation, the shape space is called Kendall's shape space (Kendall, 1984). Kendall's shape space is in general a stratified space (Goresky and MacPherson, 1988; Feragen et al., 2014), but for the special case of planar shapes, the shape space is the complex projective space, which is a complex Grassmann manifold. The examples mentioned above are often high-dimensional: a

DTI scan usually contains half a million DTs; 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, dimension reduction techniques, if applied appropriately, can benefit the subsequent statistical analysis.

For data on Riemannian manifolds, the most widely used dimensionality reduction method is the principal geodesic analysis (PGA) (Fletcher et al., 2004), which generalizes the principal component analysis (PCA) to manifold-valued data. In fact, there are many variants of PGA. Fletcher et al. (2004) proposed to find the geodesic submanifold of a certain dimension that maximizes the projected variance and computationally, it was achieved via a linear approximation, i.e., applying PCA on the tangent space at the intrinsic mean. This is sometimes referred to as the tangent PCA. Note that this approximation requires the data to be clustered around the intrinsic mean, otherwise the tangent space approximation to the manifold leads to inaccuracies. Later on, Sommer et al. (2010) proposed the Exact PGA (EPGA), which does not use any linear approximation. However, EPGA is computationally expensive as it requires 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). Chakraborty et al. (2016) partially solved this problem for manifolds with constant sectional curvature (spheres and hyperbolic spaces) by deriving closed form formulae for the projection. Other variants of PGA include but are not limited to sparse exact PGA (Banerjee et al., 2017), geodesic PCA (Huckemann et al., 2010), and probabilistic PGA (Zhang and Fletcher, 2013). All these methods focus on projecting data to a *geodesic submanifold* as in PCA where one projects data to a vector subspace. Instead, one can also project data to a submanifold that minimizes the reconstruction error without any further restrictions, e.g. being geodesic. This is the generalization of the principal curve (Hastie and Stuetzle, 1989) to Riemannian manifolds presented in Hauberg (2016).

Another feature of PCA is that it produces a sequence of nested vector subspaces. From this observation, Jung et al. (2012) proposed the *principal nested spheres* (PNS) by embedding an $n-1$-sphere into an n-sphere, and the embedding is not necessarily isometric. Hence PNS is more general than PGA in that PNS is not required to be geodesic. Similarly, for the manifold P_n of $n \times n$ SPD matrices, Harandi et al. (2018) proposed a geometry-aware dimension reduction by projecting data in P_n to P_m for some $m \ll n$. They also applied the nested P_n for the supervised dimensionality reduction problem. Damon and Marron (2014) considered a nested sequence of relations which determine a nested sequence of submanifolds that are not necessarily geodesic. They showed various examples, including Euclidean space and the n-sphere, depicting how the nested relations generalized PCA and PNS. However, for an arbitrary Riemannian manifold, it is not clear how to construct a nested submanifold. Another generalization of PGA was proposed by Pennec et al. (2018), called the exponential barycentric subspace (EBS). A k-dimensional EBS is defined as the locus of weighted exponential barycenters of $(k+1)$ affinely independent reference points. The EBSs are naturally nested by removing or adding reference points.

Unlike PGA which can be applied to any Riemannian manifolds, the construction of the nested manifolds relies heavily on the geometry of the specific manifold, and there is no general principle for such a construction. All the examples described above (spheres and P_n) and several others such as the Grassmannian, Stiefel etc. are homogeneous Riemannian manifolds (Helgason, 1979), which intuitively means that all points on the manifold 'look' the same. In this work, we propose a general framework for constructing a nested sequence of homogeneous Riemannian manifolds, and, via some simple algebraic computation, show that the nested sphere and the nested P_n can

indeed be derived within this framework. We will then apply this framework to the Grassmann manifolds, called the nested Grassmann manifolds (NG). The Grassmann manifold $\mathrm{Gr}(p, \mathbb{V})$ is the manifold of all p-dimensional subspaces of the vector space $\mathbb{V}$ where $1 \leq p \leq \dim \mathbb{V}$. Usually $\mathbb{V} = \mathbb{R}^n$ or $\mathbb{V} = \mathbb{C}^n$. An important example is Kendall's shape space of 2D shapes. The space of all shapes determined by k landmarks in $\mathbb{R}^2$ is denoted by Σ_2^k, and Kendall (1984) showed that it is isomorphic to the complex projective space $\mathbb{C}P^{k-2} \cong \mathrm{Gr}(1, \mathbb{C}^{k-1})$. In many applications, the number k of landmarks is large, and so is the dimension of $\mathrm{Gr}(1, \mathbb{C}^{k-1})$. The core of the proposed dimensionality reduction involves projecting data on $\mathrm{Gr}(p, \mathbb{V})$ to $\mathrm{Gr}(p, \widetilde{\mathbb{V}})$ with $\dim \widetilde{\mathbb{V}} \ll \dim \mathbb{V}$. The main contributions of this work are as follows: (i) We propose a general framework for constructing a nested sequence of homogeneous Riemannian manifolds unifying the recently proposed nested spheres (Jung et al., 2012) and nested SPD manifolds (Harandi et al., 2018). (ii) We present novel dimensionality reduction techniques based on the concept of NG in both supervised and unsupervised settings respectively. (iii) We demonstrate via several simulation studies and real data experiments, that the proposed NG can achieve a higher ratio of expressed variance compared to PGA.

The rest of the paper is organized as follows. In Section 2, we briefly review the definition of homogeneous Riemannian manifolds and present the recipe for the construction of nested homogeneous Riemannian manifolds. In Section 3, we first review the geometry of the Grassmannian. By applying the procedure developed in Section 2, we present the nested Grassmann manifolds representation and discuss some of its properties in details. Then we describe algorithms for our unsupervised and supervised dimensionality reduction techniques, called the Principal Nested Grassmanns (PNG), in Section 4. In Section 5, we present several simulation studies and experimental results showing how the PNG technique performs compared to PGA under different settings. Finally, we draw conclusions in Section 6.

2. Nested Homogeneous Spaces

In this section, we introduce the structure of nested homogeneous Riemannian manifolds. A Riemannian manifold (M, τ) is *homogeneous* if the group of isometries $G = I(M)$ admitted by the manifold acts transitively on M (Helgason, 1979), i.e., for $x, y \in M$, there exists $g \in G$ such that $g(x) = y$. In this case, M can be identified with G/H where H is an isotropy subgroup of G at some point $p \in M$, i.e. $H = \{g \in G : g(p) = p\}$. Examples of homogeneous Riemannian manifolds include but are not limited to, the n-spheres $S^{n-1} = \mathrm{SO}(n)/\mathrm{SO}(n-1)$, the SPD manifolds $P_n = \mathrm{GL}(n)/\mathrm{O}(n)$, the Stiefel manifolds $\mathrm{St}(m, n) = \mathrm{SO}(n)/\mathrm{SO}(n-m)$, and the Grassmann manifolds $\mathrm{Gr}(p, n) = \mathrm{SO}(n)/S(\mathrm{O}(p) \times \mathrm{O}(n-p))$.

In this paper, we focus on the case where G is either a real or a complex matrix Lie group, i.e. G is a subgroup of $\mathrm{GL}(n, \mathbb{R})$ or $\mathrm{GL}(n, \mathbb{C})$. The main idea behind the construction of nested homogeneous spaces is simple: *augmenting the matrix in G in an appropriate way*. With an embedding of the isometry group G, the embedding of the homogeneous space G/H follows naturally from the quotient structure.

Let G and $\tilde{G}$ be two connected Lie groups such that $\dim G < \dim \tilde{G}$ and $\tilde{\iota} : G \to \tilde{G}$ be an embedding. For a closed connected subgroup H of G, let $\tilde{H} = \tilde{\iota}(H)$. Since $\tilde{\iota}$ is an embedding, $\tilde{H}$ is also a closed subgroup of $\tilde{G}$. Now the canonical embedding of G/H in $\tilde{G}/\tilde{H}$ is defined by $\iota(gH) = \tilde{\iota}(g)\tilde{H}$ for $g \in G$. It is easy to see that ι is well-defined. Let $g_1, g_2 \in G$ be such that

$g_1 = g_2 h$ for some $h \in H$. Then

$$\iota(g_1 H) = \tilde{\iota}(g_1)\tilde{H} = \tilde{\iota}(g_2 h)\tilde{H} = \tilde{\iota}(g_2)\tilde{\iota}(h)\tilde{H} = \tilde{\iota}(g_2)\tilde{H} = \iota(g_2 H).$$

Now for the homogeneous Riemannian manifolds $(M = G/H, \tau_1)$ and $(\tilde{M} = \tilde{G}/\tilde{H}, \tau_2)$, denote the left-$G$-invariant, right-$H$-invariant metric on G (resp. left-$\tilde{G}$-invariant, right-$\tilde{H}$-invariant metric on $\tilde{G}$) by $\bar{\tau}_1$ and $\bar{\tau}_2$, respectively (see Cheeger and Ebin (1975, Prop. 3.16(4))).

Proposition 1 *If $\tilde{\iota}: G \to \tilde{G}$ is isometric, then so is $\iota: G/H \to \tilde{G}/\tilde{H}$.*

Proof Denote the Riemannian submersions by $\pi: G \to G/H$ and $\tilde{\pi}: \tilde{G} \to \tilde{G}/\tilde{H}$. Let X and Y be vector fields on G/H and $\bar{X}$ and $\bar{Y}$ be their horizontal lifts respectively, i.e. $d\pi(\bar{X}) = X$ and $d\pi(\bar{Y}) = Y$. By the definition of Riemannian submersions, $d\pi$ is isometric on the horizontal spaces, i.e. $\bar{\tau}_1(\bar{X}, \bar{Y}) = \tau_1(d\pi(\bar{X}), d\pi(\bar{Y})) = \tau_1(X, Y)$. Since $\tilde{\iota}$ is isometric, we have $\bar{\tau}_1(\bar{X}, \bar{Y}) = \bar{\tau}_2(d\tilde{\iota}(\bar{X}), d\tilde{\iota}(\bar{Y}))$. By the definition of ι, we also have $\iota \circ \pi = \tilde{\pi} \circ \tilde{\iota}$, which implies $d\iota \circ d\pi = d\tilde{\pi} \circ d\tilde{\iota}$. Hence,

$$\begin{aligned}
\tau_1(X, Y) = \bar{\tau}_1(\bar{X}, \bar{Y}) &= \bar{\tau}_2(d\tilde{\iota}(\bar{X}), d\tilde{\iota}(\bar{Y})) \\
&= \tau_2((d\tilde{\pi} \circ d\tilde{\iota})(\bar{X}), (d\tilde{\pi} \circ d\tilde{\iota})(\bar{Y})) \\
&= \tau_2((d\iota \circ d\pi)(\bar{X}), (d\iota \circ d\pi)(\bar{Y})) \\
&= \tau_2(d\iota(X), d\iota(Y))
\end{aligned}$$

where the third equality follows from the isometry of $d\tilde{\pi}$. ∎

Proposition 1 simply says that if the isometry group is isometrically embedded, then the associated homogeneous Riemannian manifolds will also be isometrically embedded. Conversely, if we have a Riemannian submersion $\tilde{f}: \tilde{G} \to G$, it can easily be shown that the induced map $f: \tilde{G}/\tilde{H} \to G/H$ would also be a Riemannian submersion where $H = \tilde{f}(\tilde{H})$. The construction above can be applied to a sequence of homogeneous spaces $\{M_m\}_{m=1}^{\infty}$, i.e. the embedding $\iota_m: M_m \to M_{m+1}$ can be induced from the embedding of the isometry groups $\tilde{\iota}_m: G_m \to G_{m+1}$ where $G_m = I(M_m)$ provided $\dim G_i < \dim G_j$ for $i < j$. See Figure 1 for the structure of nested homogeneous spaces.

Figure 1: Commutative diagram of the induced embedding for homogeneous spaces.

3. Nested Grassmann Manifolds

In this section, we will apply the theory of nested homogeneous space from the previous section to the Grassmann manifolds. First, we briefly review the geometry of the Grassmann manifolds in Section 3.1. With the theory in Section 2, we derive the nested Grassmann manifolds in Section 3.2, and the derivation for nested spheres and nested SPD manifolds are carried out in Section 3.3.

3.1 The Riemannian Geometry of Grassmann Manifolds

To simplify the notation, we assume $\mathbb{V} = \mathbb{R}^n$ and write $\mathrm{Gr}(p, n) := \mathrm{Gr}(p, \mathbb{R}^n)$. All the results presented in this section can be easily extended to the case of $\mathbb{V} = \mathbb{C}^n$ by replacing transposition with conjugate transposition and orthogonal groups with unitary groups. The Grassmann manifold $\mathrm{Gr}(p, n)$ is the manifold of all p-dimensional subspaces of $\mathbb{R}^n$, and for a subspace $\mathcal{X} \in \mathrm{Gr}(p, n)$, we write $\mathcal{X} = \mathrm{span}(X)$ where the columns of X form an orthonormal basis for $\mathcal{X}$. The space of all $n \times p$ matrices X such that $X^T X = I_p$ is called the *Stiefel manifold*, denoted by $\mathrm{St}(p, n)$. Special cases of Stiefel manifolds are the Lie group of all orthogonal matrices, $\mathrm{O}(n) = \mathrm{St}(n, n)$, and the n-sphere, $S^{n-1} = \mathrm{St}(1, n)$. The Stiefel manifold with the induced Euclidean metric (i.e. for $U, V \in T_X \mathrm{St}(p, n)$, $\langle U, V \rangle_X = \mathrm{tr}(U^T V)$) is a homogeneous Riemannian manifold, $\mathrm{St}(p, n) = \mathrm{SO}(n)/\mathrm{SO}(n - p)$. A canonical Riemannian metric on the Grassmann manifold can be inherited from the metric on $\mathrm{St}(p, n)$ since it is invariant to the left multiplication by elements of $O(n)$ (Absil et al., 2004; Edelman et al., 1998). The Grassmann manifold with this metric is also homogeneous, $\mathrm{Gr}(p, n) = \mathrm{SO}(n)/S(\mathrm{O}(p) \times \mathrm{O}(n - p))$.

With this canonical metric on the Grassmann manifolds, the geodesic can be expressed in closed form. Let $\mathcal{X} = \mathrm{span}(X) \in \mathrm{Gr}(p, n)$ where $X \in \mathrm{St}(p, n)$ and H be an $n \times p$ matrix. Then the geodesic $\gamma(t)$ with $\gamma(0) = \mathcal{X}$ and $\gamma'(0) = H$ is given by $\gamma_{\mathcal{X}, H}(t) = \mathrm{span}(XV \cos \Sigma t + U \sin \Sigma t)$ where $H = U \Sigma V^T$ is the compact singular value decomposition (Edelman et al., 1998, 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 $\mathrm{Exp}_{\mathcal{X}} H = \gamma_{\mathcal{X}, H}(1) = \mathrm{span}(XV \cos \Sigma + U \sin \Sigma)$. If $X^T Y$ is invertible, the geodesic distance between $\mathcal{X} = \mathrm{span}(X)$ and $\mathcal{Y} = \mathrm{span}(Y)$ is given by $d_g^2(\mathcal{X}, \mathcal{Y}) = \mathrm{tr}\, \Theta^2 = \sum_{i=1}^{p} \theta_i^2$ where $(I - XX^T)Y(X^T Y)^{-1} = U \Sigma V^T$, $U \in \mathrm{St}(p, n)$, $V \in \mathrm{O}(p)$, and $\Theta = \tan^{-1} \Sigma$. The diagonal entries $\theta_1, \ldots, \theta_k$ of Θ are known as the principal angles between $\mathcal{X}$ and $\mathcal{Y}$.
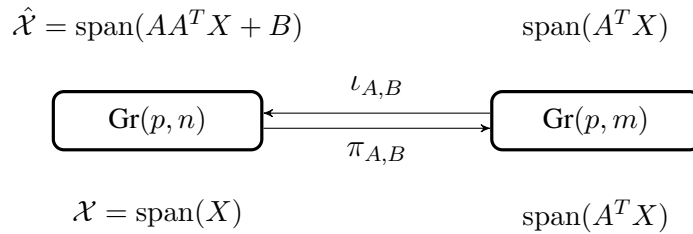
$$\hat{\mathcal{X}} = \mathrm{span}(AA^T X + B) \qquad\qquad \mathrm{span}(A^T X)$$

$$\mathrm{Gr}(p, n) \xleftarrow[\pi_{A,B}]{\iota_{A,B}} \mathrm{Gr}(p, m)$$

$$\mathcal{X} = \mathrm{span}(X) \qquad\qquad \mathrm{span}(A^T X)$$

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

3.2 Embedding of $\mathbf{Gr}(p, m)$ in $\mathbf{Gr}(p, n)$

Let $\mathcal{X} = \mathrm{span}(X) \in \mathrm{Gr}(p, m)$, $X \in \mathrm{St}(p, m)$. The map $\iota : \mathrm{Gr}(p, m) \to \mathrm{Gr}(p, n)$, for $m < n$, defined by

$$\iota(\mathcal{X}) = \mathrm{span}\left(\begin{bmatrix} X \\ 0_{(n-m)\times p} \end{bmatrix}\right)$$

is an embedding and it is easy to check that this embedding is isometric (Ye and Lim, 2016, 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 apply the framework proposed in Section 2. We consider $M_m = \mathrm{Gr}(p, m)$ for which the isometry groups are $G_m = \mathrm{SO}(m)$ and $H_m = \mathrm{S}(\mathrm{O}(p) \times \mathrm{O}(m - p))$.

In this paper, we consider the embedding $\tilde{\iota}_m : \mathrm{SO}(m) \to \mathrm{SO}(m + 1)$ given by,

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

where $O \in \mathrm{SO}(m)$, $R \in \mathrm{SO}(m + 1)$, $a, b \in \mathbb{R}^m$, $c \in \mathbb{R}$, $c \neq b^T O^{-1} a$, and $\mathrm{GS}(\cdot)$ is the Gram-Schmidt process. Since the Riemannian submersion $\pi : \mathrm{SO}(m) \to \mathrm{Gr}(p, m)$ is defined by $\pi(O) = \mathrm{span}(O_p)$ where $O \in \mathrm{SO}(m)$ and O_p is the $m \times p$ matrix containing the first p columns of O, the induced embedding $\iota_m : \mathrm{Gr}(p, m) \to \mathrm{Gr}(p, m + 1)$ is given by,

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

where $b \in \mathbb{R}^p$, $R \in \mathrm{SO}(m+1)$, $\tilde{R}$ contains the first m columns of R (which means $\tilde{R} \in \mathrm{St}(m, m+1)$), v is the last column of R, and $\mathcal{X} = \mathrm{span}(X) \in \mathrm{Gr}(p, m)$. It is easy to see that for $R = I$ and $b = 0$, this gives the natural embedding described in Ye and Lim (2016) and at the beginning of this section.

Proposition 2 *If $b = 0$, then ι_m is an isometric embedding.*

Proof With Proposition 1, it suffices to show that $\tilde{\iota}_m$ is isometric when $b = 0$. Note that as ι_m is independent of a and c in the definition of $\tilde{\iota}_m$, we can assume $a = 0$ and $c = 1$ without loss of generality. If $b = 0$, $\tilde{\iota}_m$ simplifies to

$$\tilde{\iota}_m(O) = R \begin{bmatrix} O & 0 \\ 0 & 1 \end{bmatrix}$$

where $R \in \mathrm{SO}(m + 1)$. The Riemannian distance on $\mathrm{SO}(n)$ given the induced Euclidean metric is $d_{\mathrm{SO}}(O_1, O_2) = \frac{1}{\sqrt{2}} \| \log O_1^T O_2 \|_F$. Then for $O_1, O_2 \in \mathrm{SO}(m)$,

$$d_{\mathrm{SO}}(\tilde{\iota}_m(O_1), \tilde{\iota}_m(O_2)) = \frac{1}{\sqrt{2}} \left\| \log\left(\begin{bmatrix} O_1^T O_2 & 0 \\ 0 & 1 \end{bmatrix}\right) \right\|_F = d_{\mathrm{SO}}(O_1, O_2).$$

Therefore $\tilde{\iota}_m$ is an isometric embedding, and so is ι_m by Proposition 1. ∎

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

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

Note that for $\mathcal{X} = \mathrm{span}(X) \in Gr(p, m+1)$, $\iota_m(\pi_m(\mathcal{X})) = \mathrm{span}(RR^T X + vb^T) = \mathrm{span}((I - vv^T)X + vb^T)$ where $v \in \mathbb{R}^{m+1}$ and $\|v\| = 1$. The nested relation can be extended inductively and we refer to this construction as the *nested Grassmann structure*:

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

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

Proposition 4 *The embedding of $Gr(p, m)$ into $Gr(p, n)$ for $m < n$ is given by $\iota_{A,B}(\mathcal{X}) = \mathrm{span}(AX + B)$ where $A \in St(m, n)$ and $B \in \mathbb{R}^{n \times p}$ such that $A^T B = 0$. The corresponding projection from $Gr(p, n)$ to $Gr(p, m)$ is given by $\pi_A = \mathrm{span}(A^T X)$.*

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

$$\iota_{A,B}(\mathcal{X}) = \mathrm{span}\left(\left(\prod_{i=m}^{n-1} R_i \right) X + \sum_{i=m}^{n-1} \left(\prod_{j=i+1}^{n-1} R_j \right) v_i b_i^T \right) = \mathrm{span}(AX + B)$$

where $R_i \in St(i, i + 1)$, v_i is such that $[R_i \; v_i] \in O(i + 1)$, $b_i \in \mathbb{R}^p$, $A = R_{n-1}R_{n-2} \cdots R_m \in St(m, n)$, and $B = \sum_{i=m}^{n-1} \left(\prod_{j=i+1}^{n-1} R_j \right) v_i b_i^T$ is an $n \times p$ matrix. It is easy to see that $A^T B = 0$. Similar to Proposition 3, the projection $\pi_A : Gr(p, n) \to Gr(p, m)$ is then given by $\pi_A(\mathcal{X}) = \mathrm{span}(A^T X)$. This completes the proof. ∎

From Proposition 2, if $B = 0$ then ι_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. The results discussed in this subsection can be generalized to any homogeneous space in principle. For a given homogeneous space, once the embedding of the groups of isometries (e.g., Eq. (1)) is determined, the induced embedding and the corresponding projection can be derived akin to the case of Grassmann manifolds.

3.3 Connections to Other Nested Structures

The nested homogeneous spaces proposed in this work (see Figure 1) actually provides a unified framework within which, the nested spheres (Jung et al., 2012) and the nested SPD manifolds (Harandi et al., 2018) are special cases.

The n-Sphere Example: Since the n-sphere can be identified with a homogeneous space $S^{n-1} \cong O(n)/O(n-1)$, with the embedding (1), the induced embedding of S^{n-1} into S^n is

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

where $x \in S^{n-1}$, $b \in \mathbb{R}$, and $r = \cos^{-1}\left(\frac{b}{\sqrt{1+b^2}}\right)$. This is precisely the nested sphere proposed in Jung et al. (2012, Eq. (2)).

The SPD Manifold Example: For the m-dimensional SPD manifold denoted by P_m, $G_m = \mathrm{GL}(m)$ and $H_m = O(m)$. Consider the embedding $\tilde{\iota}_m : \mathrm{GL}(m) \to \mathrm{GL}(m+1)$ given by

$$\tilde{A} = \tilde{\iota}_m(A) = R \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix} R^T,$$

where $A \in \mathrm{GL}(m)$, $R \in O(m+1)$ and the corresponding projection $\tilde{\pi}_m : \mathrm{GL}(m+1) \to \mathrm{GL}(m)$ is

$$\tilde{\pi}_m(\tilde{A}) = W^T \tilde{A} W$$

where W contains the first m columns of $R = [W \; v] \in O(m+1)$ (i.e., $W \in \mathrm{St}(m, m+1)$ and $W^T v = 0$). The submersion $\psi \circ f : \mathrm{GL}(m) \to P_m$ is given by $\psi \circ f(A) = A^T A$. Hence the induced embedding $\iota_m : P_m \to P_{m+1}$ and projection $\pi_m : P_{m+1} \to P_m$ are

$$\iota_m(X) = WXW^T + vv^T \text{ and } \pi_m(X) = W^T X W$$

which is the projection map used in Harandi et al. (2018, Eq. (13)). However, Harandi et al. (2018) did not perform any embedding or construct a nested family of SPD manifolds. Recently, it came to our attention that Jaquier and Rozo (2020) derived a similar nested family of SPD manifolds based on the projection maps in Harandi et al. (2018) described above.

4. Dimensionality Reduction with Nested Grassmanns

In this section, we discuss how to apply the nested Grassmann structure to the problem of dimension reduction. In Section 4.1 and 4.2, we describe the unsupervised and supervised dimension reduction based on the nested Grassmann manifolds. In Section 4.3, we will discuss the choice of distance metrics required by the dimensionality reduction algorithm and present some technical details regarding the implementation. Analysis of principal nested Grassmann (PNG) will be introduced and discussed in Section 4.4 and Section 4.5.

4.1 Unsupervised Dimensionality Reduction

We can now apply the nested Grassmann structure to the problem of unsupervised dimensionality reduction. Suppose that we are given the points, $\mathcal{X}_1, \dots, \mathcal{X}_N \in \mathrm{Gr}(p, n)$. We would like to have lower dimensional representations in $\mathrm{Gr}(p, m)$ for $\mathcal{X}_1, \dots, \mathcal{X}_N$ with $m \ll n$. The desired projection map π_A that we seek is obtained by the minimizing the reconstruction error, i.e.

$$L_u(A, B) = \frac{1}{N} \sum_{i=1}^{N} d^2(\mathcal{X}_i, \hat{\mathcal{X}}_i) = \frac{1}{N} \sum_{i=1}^{N} d^2(\mathrm{span}(X_i), \mathrm{span}(AA^T X_i + B))$$

where $d(\cdot, \cdot)$ 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. $L_u(AO, B) = L_u(A, B)$ for $O \in \mathrm{O}(m)$. 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 (Edelman et al., 1998) to obtain A and B by optimizing $L_u(A, B)$ over $\mathrm{span}(A) \in \mathrm{Gr}(m, n)$ and $B \in \mathbb{R}^{n \times p}$ such that $A^T B = 0$. Note that the restriction $A^T B = 0$ simply means that the columns of B are in the null space of A^T, denoted $N(A^T)$. Hence in practice this restriction can be handled as follows. For arbitrary $\tilde{B} \in \mathbb{R}^{n \times p}$, project $\tilde{B}$ on to $N(A^T)$, i.e. $B = P_{N(A^T)} \tilde{B}$ where $P_{N(A^T)} = I - AA^T$ is the projection from $\mathbb{R}^n$ to $N(A^T)$. Thus, the loss function can be written as

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

and it is optimized over $\mathrm{Gr}(m, n) \times \mathbb{R}^{n \times p}$. Note that since we represent a subspace by its orthonormal basis, when $m > n/2$, we can use the isomorphism $\mathrm{Gr}(m, n) \cong \mathrm{Gr}(n - m, n)$ to further reduce the computational burden. This will be particularly useful when $m = n - 1$ as in Section 4.4. Under this isomorphism $\mathrm{Gr}(m, n) \cong \mathrm{Gr}(n - m, n)$, the corresponding subspace of $\mathrm{span}(A) \in \mathrm{Gr}(m, n)$ is $\mathrm{span}(A_\perp) \in \mathrm{Gr}(n - m, n)$ where $A_\perp$ is an $n \times (n - m)$ matrix such that $[A \; A_\perp]$ is an orthogonal matrix. Hence the loss function L_u can alternatively be expressed as

$$L_u(A, B) = \frac{1}{N} \sum_{i=1}^{N} d^2(\mathrm{span}(X_i), \mathrm{span}((I - A_\perp A_\perp^T)X_i + A_\perp A_\perp^T B)).$$

Remark 1 *The reduction of the space of all possible projections from $St(m, n)$ to $Gr(m, n)$ is a consequence of the choice of the loss function L_u. With a different loss function, one might have a different space of possible projections.*

4.2 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 use this extra information to sharpen the result of dimensionality reduction. Specifically, we expect that after reducing the dimension, points from the same class preserve their proximity while points from different classes are well separated. We use an *affinity function* $a : \mathrm{Gr}(p, n) \times \mathrm{Gr}(p, n) \rightarrow \mathbb{R}$ to encode the structure of the data as suggested by Harandi et al. (2018, Sec 3.1, Eq. (14)-(16)). For completeness, we repeat the definition of the affinity function here. The affinity function is defined as $a(\mathcal{X}_i, \mathcal{X}_j) = g_w(\mathcal{X}_i, \mathcal{X}_j) - g_b(\mathcal{X}_i, \mathcal{X}_j)$ where

$$g_w(\mathcal{X}_i, \mathcal{X}_j) = \begin{cases} 1 & \text{if } \mathcal{X}_i \in N_w(\mathcal{X}_j) \text{ or } \mathcal{X}_j \in N_w(\mathcal{X}_i) \\ 0 & \text{Otherwise} \end{cases}$$

$$g_b(\mathcal{X}_i, \mathcal{X}_j) = \begin{cases} 1 & \text{if } \mathcal{X}_i \in N_b(\mathcal{X}_j) \text{ or } \mathcal{X}_j \in N_b(\mathcal{X}_i) \\ 0 & \text{Otherwise} \end{cases}.$$

The set $N_w(\mathcal{X}_i)$ consists of ν_w nearest neighbors of $\mathcal{X}_i$ that have the *same* labels as y_i, and the set $N_b(\mathcal{X}_i)$ consists of ν_b nearest neighbors of $\mathcal{X}_i$ that have *different* labels from y_i. The nearest neighbors can be computed using the geodesic distance.

The desired projection map π_A that we seek is obtained by the minimizing the following loss function

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

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

4.3 Choice of the distance function

The loss functions L_u and L_s depend on the choice of the distance $d : \text{Gr}(p, n) \times \text{Gr}(p, n) \to \mathbb{R}_{\geq 0}$. Besides the geodesic distance, there are many widely used distances on the Grassmann manifold, see, for example, Edelman et al. (1998, p. 337) and Ye and Lim (2016, 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 Ye and Lim (2016) and the projection F-norm in Edelman et al. (1998). The geodesic distance was defined in Section 3.1 and the projection distance is defined as follows. For $\mathcal{X}, \mathcal{Y} \in \text{Gr}(p, n)$, 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 $\mathcal{X} = \text{span}(X)$, then $P_{\mathcal{X}} = X(X^T X)^{-1} X^T$. It is also easy to see the the projection distance has $O(n)$-symmetry. We choose the projection distance mainly for its computational efficiency as it involves only matrix multiplication which has a time complexity $O(n^2)$ whereas the geodesic distance requires an SVD which has a time complexity of $O(n^3)$.
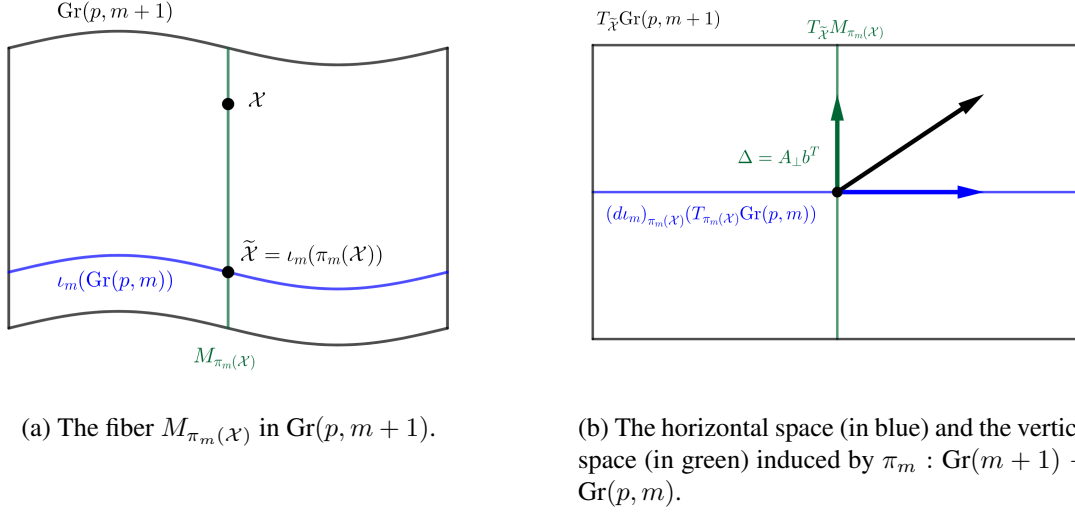
4.4 Analysis of Principal Nested Grassmannians

To determine the dimension of the nested submanifold that fits the data well enough, we can choose $p < m_1 < \ldots < m_k < n$ and estimate the projection onto these nested Grassmann manifolds. The ratio of expressed variance for each projection is the ratio of the variance of the projected data and the variance of the original data. With these ratios, we can choose the desired dimension according to the pre-specified percentage of expressed variance as one would do for choosing the number of principal components in PCA.

Alternatively, one can have a full analysis of principal nested Grassmanns (PNG) as follows. Starting from $\text{Gr}(p, n)$, one can reduce the dimension down to $\text{Gr}(p, p + 1)$. Using the diffeomorphism between $\text{Gr}(p, n)$ and $\text{Gr}(p, n - p)$, we have $\text{Gr}(p, p + 1) \cong \text{Gr}(1, p + 1)$, and hence we can continue reducing the dimension down to $\text{Gr}(1, 2)$. The resulting sequence will be

$$\text{Gr}(p, n) \to \text{Gr}(p, n - 1) \to \cdots \to \text{Gr}(p, p + 1) = \text{Gr}(1, p + 1) \to \text{Gr}(1, p) \to \cdots \to \text{Gr}(1, 2).$$

Furthermore, we can reduce the points on $\text{Gr}(1, 2)$, which is a 1-dimensional manifold, to a 0-dimensional manifold, which is simply a point, by computing the Fréchet mean (FM). We call this FM the nested Grassmannian mean (NGM) of $\mathcal{X}_1, \ldots, \mathcal{X}_N \in \text{Gr}(p, n)$. The NGM is unique since $\text{Gr}(1, 2) \cong \mathbb{R}P^1$ can be identified as the half circle in $\mathbb{R}^2$ and the FM is unique in this case. Note that in general, the NGM will not be the same as the FM of $\mathcal{X}_1, \ldots, \mathcal{X}_N$ since the embedding/projection

(a) The fiber $M_{\pi_m(\mathcal{X})}$ in $\mathrm{Gr}(p, m+1)$.

(b) The horizontal space (in blue) and the vertical space (in green) induced by $\pi_m : \mathrm{Gr}(m+1) \to \mathrm{Gr}(p, m)$.

Figure 3: Illustrations of submanifolds induced by π_m.

need not be isometric. The supervised PNG (sPNG) can be obtained similarly by replacing each projection with it supervised counterpart.

4.5 Principal Scores

Whenever we apply a projection $\pi_m : \mathrm{Gr}(p, m+1) \to \mathrm{Gr}(p, m)$ to the data, we might lose some information contained in the data. More specifically, since we project data on a $p(m+1-p)$-dimensional manifold to a $p(m-p)$-dimensional manifold, we need to describe this $p(m+1-p) - p(m-p) = p$ dimensional information loss during the projection. In PCA, this is done by computing the scores of each principal component, which are the transformed coordinates of each sample in the eigenspace of the covariance matrix. We can generalize the notion of principal scores to the nested Grassmanns as follows: For each $\mathcal{X} \in \mathrm{Gr}(p, m+1)$, denote by $M_{\pi_m(\mathcal{X})}$, the fiber of $\pi_m(\mathcal{X})$, i.e. $M_{\pi_m(\mathcal{X})} = \pi_m^{-1}(\pi_m(\mathcal{X})) = \{\mathcal{Y} \in \mathrm{Gr}(p, m+1) : \pi_m(\mathcal{Y}) = \pi_m(\mathcal{X})\}$, which is a p-dimensional submanifold of $\mathrm{Gr}(p, m+1)$. An illustration of this fiber is given in Figure 3a. Let $\widetilde{\mathcal{X}} = \iota_m(\pi_m(\mathcal{X}))$ and let the unit tangent vector $V \in T_{\widetilde{\mathcal{X}}} M_{\pi_m(\mathcal{X})}$ be the geodesic direction from $\widetilde{\mathcal{X}}$ to $\mathcal{X}$. Given a suitable basis on $T_{\widetilde{\mathcal{X}}} M_{\pi_m(\mathcal{X})}$, V can be realized as a p-dimensional vector, and this will be the score vector of $\mathcal{X}$ associated with the projection π_m.

By the definition of $M_{\pi_m(\mathcal{X})}$, we have the following decomposition of the tangent space of $\mathrm{Gr}(p, m+1)$ at $\widetilde{\mathcal{X}}$ into the horizontal space and the vertical space induced by π_m,

$$T_{\widetilde{\mathcal{X}}} \mathrm{Gr}(p, m+1) = T_{\widetilde{\mathcal{X}}} M_{\pi_m(\mathcal{X})} \oplus (d\iota_m)_{\pi_m(\mathcal{X})}(T_{\pi_m(\mathcal{X})} \mathrm{Gr}(p, m)).$$

An illustration of this decomposition is given in Figure 3b. A tangent vector of $M_{\pi_m(\mathcal{X})}$ at $\widetilde{\mathcal{X}}$ is of the form $\Delta = A_\perp b^T$ where $A_\perp$ is any $(m+1)$-dim vector such that $[A \ A_\perp]$ is orthogonal and $b \in \mathbb{R}^p$. It is easy to check that $\pi_m(\mathrm{span}(AA^T X + A_\perp b^T)) = \pi_m(\mathrm{span}(X)) = \mathrm{span}(A^T X)$. Hence a natural coordinate for the tangent vector $\Delta = A_\perp b^T$ is $b \in \mathbb{R}^p$, and the geodesic direction from $\widetilde{\mathcal{X}}$ to $\mathcal{X}$ would be $V = X^T A_\perp$. It is easy to see that $\|V\|_F = 1$ since X has orthonormal columns. To reflect the distance between $\widetilde{\mathcal{X}}$ and $\mathcal{X}$, i.e. the reconstruction error, we define $d(\widetilde{\mathcal{X}}, \mathcal{X})V$ as the

score vector for $\mathcal{X}$ associated with π_m. In the case of $\mathrm{Gr}(1,2) \to$ NGM, we use the sign distance to the NGM as the score. For complex nested Grassmanns however, the principal score associated with each projection is a p-dimensional complex vector. For the sake of visualization, we transform this p-dimensional complex vector to a $2p$-dimensional real vector. The procedure for computing the PNG and the principal scores is summarized in Algorithm 1.

Remark 2 *Note that this definition of principal score is not intrinsic as it depends on the choice of basis. Indeed, it is impossible to choose a p-dimensional vector for the projection π_m in an intrinsic way, since the only property of a map that is independent of bases is the rank of the map. A reasonable choice of basis is made by viewing the Grassmann $Gr(p, m)$ as a quotient manifold of $St(p, m)$, which is a submanifold in $\mathbb{R}^{m \times p}$. This is how we define the principal score for nested Grassmanns.*

5. Experiments

In this section, we will demonstrate the performance of the proposed dimensionality reduction technique, i.e. PNG and sPNG, via experiments on synthetic and real data. The implementation[1] is based on the python library `pymanopt` (Townsend et al., 2016) and we use the steepest 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.

5.1 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 4.3, 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?

5.1.1 PROJECTION AND GEODESIC DISTANCE COMPARISONS

The procedure we used to generate random points on $\mathrm{Gr}(p, n)$ for the synthetic data experiments is as follows: First, we generate N points from a uniform distribution on $\mathrm{St}(p, m)$ (Chikuse, 2003, 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 $\widetilde{\mathcal{X}}_i = \mathrm{span}(AX_i + (I - AA^T)B) \in \mathrm{Gr}(p, n)$. Finally, we compute $\mathcal{X}_i = \mathrm{Exp}_{\widetilde{\mathcal{X}}_i}(\sigma U_i)$, where $U_i = \widetilde{U}_i / \|\widetilde{U}_i\|$ and $\widetilde{U}_i \in T_{\widetilde{\mathcal{X}}_i}\mathrm{Gr}(p, n)$, to include some perturbation.

This experiment involves comparing the performance of the NG representation in terms of the explained variance, 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 0.5 to 1. The results are averaged over 100 repetitions and are shown in Figure 4. From these results, we can see that the explained variance for the projection distance and the geodesic distance are indistinguishable but using projection distance leads to much faster convergence than when using the geodesic distance. The reason is that when

---

1. Our code is available at `https://github.com/cvgmi/NestedGrassmann`.

Algorithm 1: Principal Nested Grassmanns

Input : $\mathcal{X}_1, \ldots \mathcal{X}_N \in \text{Gr}(p, n)$
Output: An $N \times p(n-p)$ score matrix.
Let $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_N)$.
/* $\text{Gr}(p, n) \to \text{Gr}(p, n-1) \to \cdots \to \text{Gr}(p, p+1)$ */
for $m = n-1, \ldots, p+1$ **do**
$\quad \hat{v}, \hat{b} = \arg\min_{v,b} \sum_{i=1}^N d^2(\mathcal{X}_i, \text{span}((I - vv^T)X_i + vb^T))$;
$\quad$ /* $v \in \mathbb{R}^{m+1}$, $\|v\| = 1$, $b \in \mathbb{R}^p$ */
$\quad$ **for** $i = 1, \ldots, N$ **do**
$\quad\quad$ Let $\widehat{\mathcal{X}}_i = \text{span}((I - \hat{v}\hat{v}^T)X_i + \hat{v}\hat{b}^T)$ where $\mathcal{X}_i = \text{span}(X_i)$.
$\quad\quad V_{i,m} = d_g(\mathcal{X}_i, \widehat{\mathcal{X}}_i)X_i^T v$;   // the score vector for $\mathcal{X}_i$
$\quad\quad \mathcal{X}_i \leftarrow \text{span}(R^T X_i)$;   // $R \in \text{St}(m, m+1)$ such that $R^T \hat{v} = 0$.

if $p > 1$ **then**
$\quad$ **for** $i = 1, \ldots, N$ **do**
$\quad\quad \mathcal{X}_i \leftarrow \mathcal{X}_i^\perp$;   // $\text{Gr}(p, p+1) \cong \text{Gr}(1, p+1)$
$\quad$ /* $\text{Gr}(1, p+1) \to \text{Gr}(1, p) \to \cdots \to \text{Gr}(1, 2)$ */
$\quad$ **for** $m = p, \ldots, 2$ **do**
$\quad\quad \hat{v}, \hat{b} = \arg\min_{v,b} \sum_{i=1}^N d^2(\mathcal{X}_i, \text{span}((I - vv^T)X_i + vb^T))$;
$\quad\quad$ /* $v \in \mathbb{R}^{m+1}$, $\|v\| = 1$, $b \in \mathbb{R}^p$ */
$\quad\quad$ **for** $i = 1, \ldots, N$ **do**
$\quad\quad\quad$ Let $\widehat{\mathcal{X}}_i = \text{span}((I - \hat{v}\hat{v}^T)X_i + \hat{v}\hat{b}^T)$ where $\mathcal{X}_i = \text{span}(X_i)$.
$\quad\quad\quad V_{i,m} = d_g(\mathcal{X}_i, \widehat{\mathcal{X}}_i)X_i^T v$;   // the score vector for $\mathcal{X}_i$
$\quad\quad\quad \mathcal{X}_i \leftarrow \text{span}(R^T X_i)$;   // $R \in \text{St}(m, m+1)$ such that $R^T v = 0$.

/* $\text{Gr}(1, 2) \to \text{NGM}$ */
/* Note that $\widehat{\mathcal{X}}_i \in \text{Gr}(1, 2)$ */
$\text{NGM} \leftarrow \text{FM}(\widehat{\mathcal{X}}_1, \ldots, \widehat{\mathcal{X}}_N)$;
Let $U_i \in \mathbb{R}^2$ be the geodesic direction from the NGM to $\mathcal{X}_i$ such that
$\|U_i\| = d_g(\text{NGM}, \mathcal{X}_i)$.;
for $i = 1, \ldots, N$ **do**
$\quad V_{i,1} \in \mathbb{R}$ is such that $U_i = V_{i,1} \times \frac{U_1}{\|U_1\|}$;
Return the score matrix $S = [V_{i,m}]$. (Note that $V_{i,m} \in \mathbb{R}$ for $m = 1, \ldots, p$ and $V_{i,m} \in \mathbb{R}^p$
for $m = p+1, \ldots, n-1$.)

---

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 all the experiments in the next section, we use the projection distance for the sake of efficiency.
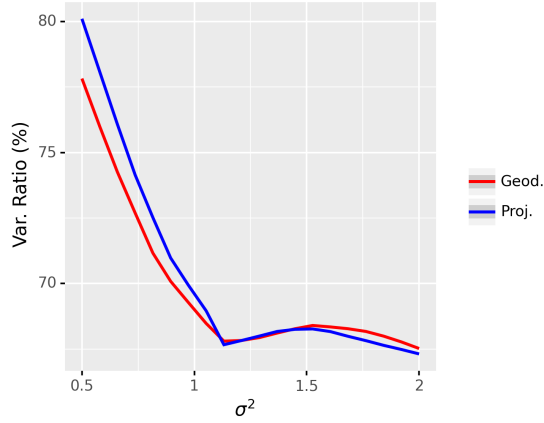

Figure 4: Comparison of the NG representations based on the projection and geodesic distances using the expressed variance.

5.1.2 COMPARISON OF PNG AND PGA

Now we compare the proposed PNG to PGA. In order to have a fair comparison between PNG and PGA, we define the principal components of PNG as the principal components of the scores obtained as in Section 4.5. Similar to the previous experiment, we set $N = 50$, $n = 10$, $m = 5$, $p = 2$, and $\sigma = 0.01, 0.05, 0.1, 0.5$ and apply the same procedure to generate synthetic data. The results are averaged over 100 repetitions and are shown in Figure 5.
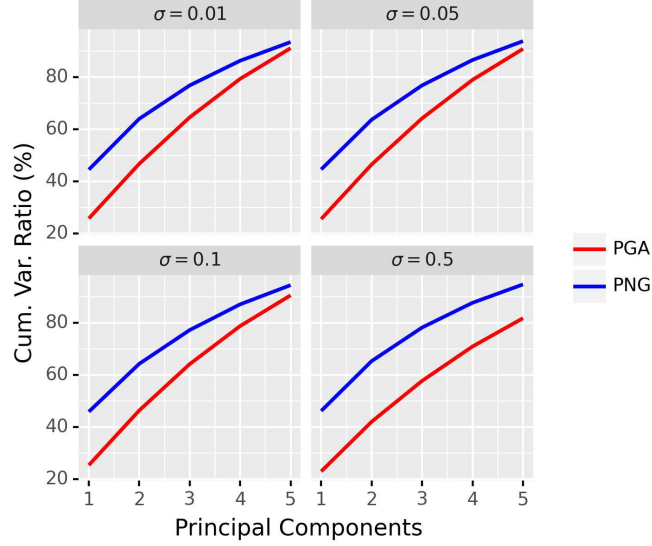
14

Figure 5: Comparison of PNG and PGA under different levels of noise.

From Figure 5, 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. We can see that when the variance is small, the improvement of PNG over PGA is less significant, whereas, our method is significantly better for the large data variance case (e.g. comparing $\sigma = 0.5$ and $\sigma = 0.01$). 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 (Jung et al., 2012) 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.

5.2 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 a k-ad or a configuration. Here we assume that these k points are not all identical. Denote the configuration by X which is a $k \times 2$ matrix. Let H be the $(k-1) \times k$ Helmert submatrix (Dryden and Mardia, 2016, Ch. 2.5). Then $Z = HX/\|HX\|_F$ is called the *pre-shape* of X from which the information about translation and scaling is removed. The space of all pre-shapes is called the pre-shape space, denoted by $\mathcal{S}_2^k$. By definition, the pre-shape space is a $(2k-3)$-dimensional sphere. The shape is obtained by removing the rotation from the pre-shape, and thus the shape space is $\Sigma_2^k = \mathcal{S}_2^k/\mathrm{O}(2)$. It was shown by Kendall (1984) that Σ_2^k is a smooth manifold and, when equipped with the quotient metric, is isometric to the complex projective space $\mathbb{C}P^{k-2}$ equipped with the Fubini-Study metric (up to a scale factor)

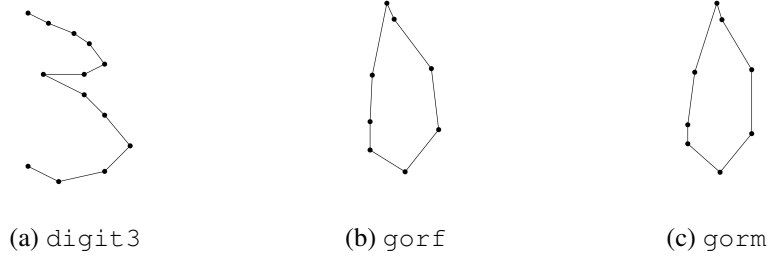(a) `digit3` (b) `gorf` (c) `gorm`

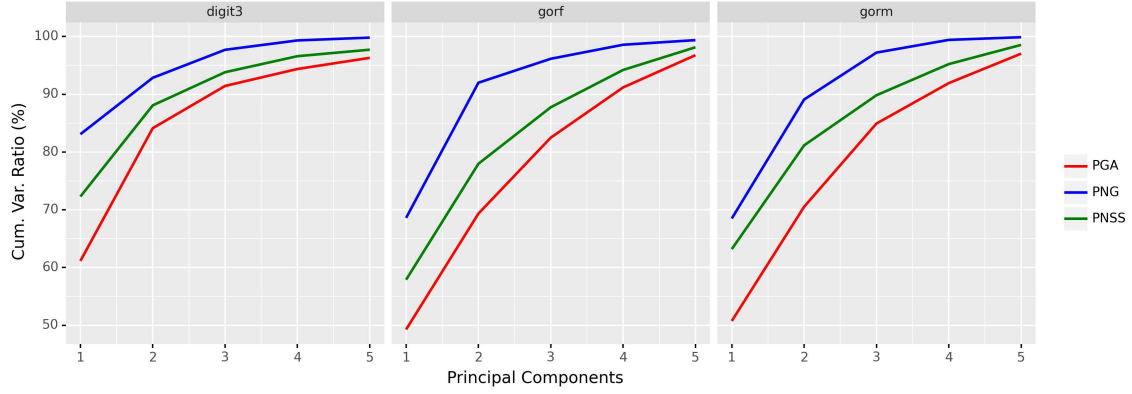Figure 6: Example shapes from the three datasets.

Figure 7: Cumulative explained variance by the first 5 principal components of PNG, PGA, and PNSS.

which is a special case of the complex Grassmannians, i.e. $\mathbb{C}P^{k-2} \cong \mathrm{Gr}(1, \mathbb{C}^{k-1})$. Hence, we can apply the proposed PNG to planar shapes. For planar shapes, we also compare with the recently proposed principal nested shape spaces (PNSS) (Dryden et al., 2019), which is an application of PNS on the pre-shape space. We will now demonstrate how the PNG performs compared to PGA and PNSS using some simple examples of planar shapes and the OASIS dataset.

Examples of Planar Shapes We take three datasets: `digit3`, `gorf`, and `gorm`, from the R package `shapes` (Dryden, 2021). The `digit3` dataset consists of 30 shapes of the digit 3, each of which is represented by 13 points in $\mathbb{R}^2$; the `gorf` dataset consists of 30 shapes of female gorilla skull, each of which is represented by 8 points in $\mathbb{R}^2$; the `gorm` dataset consists of 29 shapes of male gorilla skull, each of which is represented by 8 points in $\mathbb{R}^2$. Example shapes from these three datasets are shown in Figure 6. The cumulative ratios of variance explained by the first 5 principal components[2] of PNG, PGA, and PNSS are shown in Figure 7. It can be seen from Figure 7 that the proposed PNG achieves higher explained variance than PGA and PNSS respectively in most cases.

OASIS Corpus Callosum Data Experiment The OASIS database (Marcus et al., 2007) is a publicly available database that contains T1-MR brain scans of subjects of age ranging from 18 to

---

2. Here the principal components in PNG and PGA are complex whereas the principal components in PNSS are real. Hence, we choose 5 principal components in PNG and PGA and 10 principal components in PNSS, so that the reduced (real) dimension is 10 in all three cases.

96. In particular, it includes subjects that are clinically diagnosed with mild to moderate Alzheimer's 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 8 for samples of the segmented corpus callosi. In this case, the shape space is $\Sigma_2^{248} \cong \mathbb{C}P^{248} \cong \mathrm{Gr}(1, \mathbb{C}^{249})$. Results of application of the three methods to this data are shown in Figure 9.
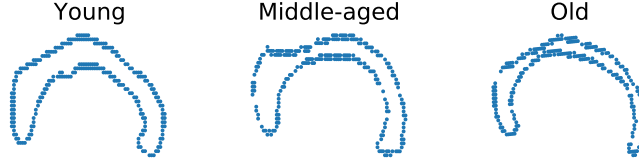

Figure 8: Example Corpus Callosi shapes from three distinct age groups, each depicted using the boundary point sets.
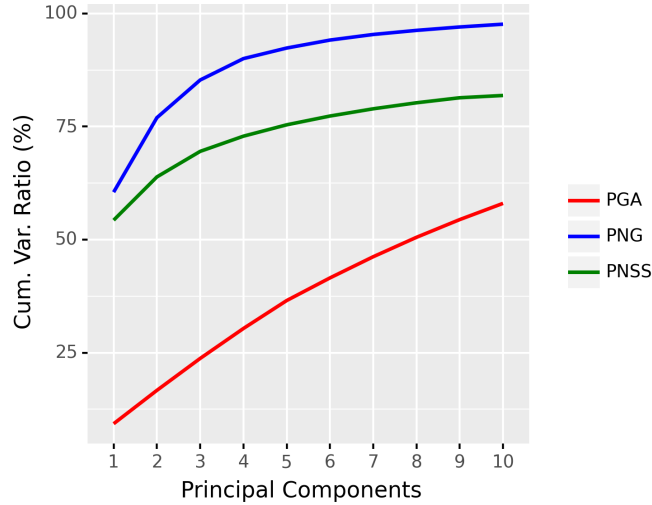

Figure 9: Cumulative explained variance captured by the first 10 principal components of PNG, PGA, and PNSS respectively.

Since the data are divided into three groups (young, middle-aged, and old), we can apply the sPNG described in Section 4.2 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 simple classifier such as the support vector machine (SVM) Vapnik (1995) 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 Bair et al. (2006); Barshan et al. (2011) on the tangent space. However, generalizing PNSS to the supervised case is nontrivial and is beyond the scope of this paper. Therefore, we limit our comparison to the unsupervised PNSS. In this demonstration, we apply an SVM on the scores obtained from different dimension reduction algorithms, and we choose only the first three principal scores to show that even with the 3-dimensional representation of the original shapes, we can still achieve good classification results. The results are shown in Table 1. These results are in accordance with our expectation since in sPNG, 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 sPNG exhibits far superior performance compared to sPGA in accuracy.

Method	Accuracy
sPNG	83.33%
PNG	75%
sPGA	66.67%
PGA	63.89%
PNSS	80.56%

Table 1: Classification accuracies for sPGA and sPNG respectively.

6. Conclusion

In this work, we proposed a novel nested geometric structure for homogeneous spaces and used this structure to achieve dimensionality reduction for data residing in Grassmann manifolds. We also discuss how this nested geometric structure served as a natural generalization of other existing nested geometric structures in literature namely, 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 submanifold to be a geodesic submanifold and to pass through the Fréchet mean of the data. Succinctly, 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 and the recently proposed PNSS by a large margin.

Acknowledgments

Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects.

Conflicts of Interest

We declare we don't have conflicts of interest.

References

P-A Absil, Robert Mahony, and Rodolphe Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematica*, 80(2):199–220, 2004.

Eric Bair, Trevor Hastie, Debashis Paul, and Robert Tibshirani. Prediction by supervised principal components. *Journal of the American Statistical Association*, 101(473):119–137, 2006.

Monami Banerjee, Rudrasis Chakraborty, and Baba C Vemuri. Sparse exact PGA on Riemannian manifolds. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5010–5018, 2017.

Elnaz Barshan, Ali Ghodsi, Zohreh Azimifar, and Mansoor Zolghadri Jahromi. Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. *Pattern Recognition*, 44(7):1357–1371, 2011.

Peter J Basser, James Mattiello, and Denis LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66(1):259–267, 1994.

Paul T Callaghan. *Principles of Nuclear Magnetic Resonance Microscopy*. Oxford University Press on Demand, 1993.

Rudrasis Chakraborty, Dohyung Seo, and Baba C Vemuri. An efficient exact-PGA algorithm for constant curvature manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3976–3984, 2016.

Jeff Cheeger and David Gregory Ebin. *Comparison theorems in Riemannian geometry*, volume 9. North-holland Amsterdam, 1975.

Yasuko Chikuse. *Statistics on Special Manifolds*, volume 174. Springer Science & Business Media, 2003.

James Damon and JS Marron. Backwards principal component analysis and principal nested relations. *Journal of Mathematical Imaging and Vision*, 50(1-2):107–114, 2014.

I. L. Dryden. `shapes` *package*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL `http://www.R-project.org`. Contributed package, Version 1.2.6.

Ian L Dryden and Kanti V Mardia. *Statistical shape analysis: with applications in R*, volume 995. John Wiley & Sons, 2016.

Ian L Dryden, Kwang-Rae Kim, Charles A Laughton, Huiling Le, et al. Principal nested shape space analysis of molecular dynamics data. *Annals of Applied Statistics*, 13(4):2213–2234, 2019.

Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

Aasa Feragen, Mads Nielsen, Eva Bjørn Vedel Jensen, Andrew du Plessis, and François Lauze. Geometry and statistics: Manifolds and stratified spaces. *Journal of Mathematical Imaging and Vision*, 50(1):1–4, 2014.

P Thomas Fletcher, Conglin Lu, Stephen M Pizer, and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8): 995–1005, 2004.

Mark Goresky and Robert MacPherson. *Stratified Morse Theory*. Springer, 1988.

Mehrtash Harandi, Mathieu Salzmann, and Richard Hartley. Dimensionality reduction on SPD manifolds: The emergence of geometry-aware methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(1):48–62, 2018.

Trevor Hastie and Werner Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84(406):502–516, 1989.

Søren Hauberg. Principal curves on Riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(9):1915–1921, 2016.

Sigurdur Helgason. *Differential geometry, Lie groups, and symmetric spaces*. Academic press, 1979.

Stephan Huckemann, Thomas Hotz, and Axel Munk. Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statistica Sinica*, pages 1–58, 2010.

Noémie Jaquier and Leonel Rozo. High-dimensional bayesian optimization via nested riemannian manifolds. *Advances in Neural Information Processing Systems*, 33:20939–20951, 2020.

Sungkyu Jung, Ian L Dryden, and JS Marron. Analysis of principal nested spheres. *Biometrika*, 99 (3):551–568, 2012.

David G Kendall. Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. 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.

Xavier Pennec et al. Barycentric subspace analysis on manifolds. *The Annals of Statistics*, 46(6A): 2711–2746, 2018.

Stefan Sommer, François Lauze, Søren Hauberg, and Mads Nielsen. Manifold valued statistics, exact principal geodesic analysis and the effect of linear approximations. In *European Conference on Computer Vision*, pages 43–56. Springer, 2010.

James Townsend, Niklas Koep, and Sebastian Weichwald. Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation. *Journal of Machine Learning Research*, 17(137):1–5, 2016. URL `http://jmlr.org/papers/v17/16-177.html`.

Vladimir Vapnik. *The Nature of Statistical Learning Theory*. Springer Science & Business Media, 1995.

Ke Ye and Lek-Heng Lim. Schubert varieties and distances between subspaces of different dimensions. *SIAM Journal on Matrix Analysis and Applications*, 37(3):1176–1197, 2016.

Miaomiao Zhang and Tom Fletcher. Probabilistic principal geodesic analysis. In *Advances in Neural Information Processing Systems*, pages 1178–1186, 2013.