HAL
archives-ouvertes.fr

Barycentric Subspaces Analysis on Spheres

Xavier Pennec

▶ **To cite this version:**

Xavier Pennec. Barycentric Subspaces Analysis on Spheres. Mathematical Foundation sof Computational Anatomy (MFCA'15), Oct 2015, Munich, Germany. pp.71-82, 2015, Proceedings of the fifth international workshop on Mathematical Foundation sof Computational Anatomy (MFCA'15). <hal-01203815>

HAL Id: hal-01203815
https://hal.inria.fr/hal-01203815

Submitted on 23 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Barycentric Subspaces Analysis on Spheres

Xavier Pennec[1]

Inria Sophia-Antipolis and Côte d'Azur University (UCA), France
xavier.pennec@inria.fr

Abstract. This paper addresses the generalization of Principal Component Analysis (PCA) to Riemannian manifolds. Current methods like Principal Geodesic Analysis (PGA) and Geodesic PCA (GPCA) minimize the distance to a "Geodesic subspace". This allows to build sequences of nested subspaces which are consistent with a forward component analysis approach. However, these methods cannot easily be adapted to a backward analysis and alck symmetric in the parametrization of the subspaces. We propose in [10] a new and more general type of family of subspaces in manifolds, *barycentric subspaces*, which are implicitly defined as the locus of points which are weighted means of $k+1$ reference points. Depending on the generalization of the mean that we use, we obtain the Fréchet / Karcher / Exponential barycentric subspaces (FBS / KBS / EBS). The completion of the last one is called the affine span. These definitions were shown to define locally submanifolds of dimension k.
In this paper, we investigate barycentric subspaces in one of the simplest manifold: the sphere. We show that the affine span is a great subsphere in generic conditions, i.e. also a geodesic subspace. This coincidence of spaces is due to the very high symmetry of the sphere. For second order jets, we show that we obtain subspheres of different radii as in the case of principal nested spheres (PNS) analysis. Among the points of the affine span, determining which ones belong to the Karcher barycentric subspaces (KBS) turns out to be a surprisingly difficult problem. Practical experiments show that the KBS covers in general only a small portion of the subsphere containing the reference points. This suggests that the affine span might be a much more interesting definition to work with for subspace definition purposes. We finally discuss the use to these barycentric subspaces to generalize PCA on manifolds, a procedure that we name Barycentric Subpsace Analysis (BSA). Like PGA, barycentric subspaces allow the construction of a forward nested sequence of subspaces which contains the Fréchet mean. However, the definition also allows the construction of backward nested sequence which may not contain the mean.

1 Introduction

In a Euclidean space, the principal k-dimensional affine subspace of the Principal Component Analysis (PCA) procedure is equivalently defined by minimizing the variance of the residuals (the projection of the data point to the subspace) or by maximizing the explained variance within that affine subspace. This is due to

Pythagoras' theorem, which does not hold in more general manifolds. A second important observation is that principal components or different orders are nested, which allows to build forward and backward estimation methods by iteratively adding or removing principal components.

Generalizing PCA to manifolds first requires to define the equivalent of affine subspaces in manifolds. For the zero-dimensional subspace, intrinsic generalization of the mean on manifolds naturally comes into mind: the Fréchet mean is the set of global minima of the variance, as defined by Fréchet in general metric spaces [3]. The set of local minima of the variance was named Karcher mean by W.S Kendall [8] after the work of Karcher et al. [6] on Riemannian centers of mass (see [7] for a discussion of the naming and earlier works). The one-dimensional component is then quite naturally a geodesic which should pass through the mean point. Higher-order components are more difficult to define. The simplest generalization is tangent PCA (tPCA), which amounts to unfold the whole distribution in the tangent space at the mean, and to compute the principal components of the covariance matrix in the tangent space. The method is thus based on the maximization of the explained variance. tPCA was used implicitly or explicitly in a lot of statistical works on shape spaces and Riemannian manifolds because it is simple and efficient. However, if tPCA is good for analyzing data which are sufficiently centered around a central value (unimodal or Gaussian-like data), it is often not sufficient for multimodal or large support distributions (e.g. uniform on close compact subspaces).

Instead of an analysis of the covariance matrix, Fletcher et al. [2] proposed to rely on the least square distance to subspaces which are totally geodesic at a point. These *Geodesic Subspaces (GS)* are spanned by the geodesics going through a point with tangent vector restricted to belong to a linear subspace of the tangent space. The procedure was coined Principal Geodesic Analysis (PGA). However, the least-square procedure was computationally expensive, so that it was approximated in practice with tPCA. A complete implementation of the original PGA procedure was only provided recently by Sommer et al. [14]. PGA is intrinsic and allows to build a flag (sequences of embedded subspaces) of principal geodesic subspaces which is consistent with a forward component analysis approach: we build iteratively the components from dimension 0 (the mean point), dimension 1 (a geodesic) and higher dimensions by iteratively selecting the direction in the tangent space at the mean that optimally reduce the square distance of data point to the geodesic subspace. In this procedure, the mean always belong to geodesic subspaces even when they are not part of the support of the distribution.

To alleviate this problem, Huckemann et al. [12] proposed to relax the fact that the base-point of the geodesic subspace has to be the Fréchet mean: they start at the first order component directly with the geodesic that best fits the data, which is not necessarily going through the mean. The second principal geodesic is chosen orthogonally to the first one, and higher order components are added orthogonally at the crossing point of the first two components. The method was named Geodesic PCA (GPCA). Further relaxing the assumption

that second and higher order components should cross at a single point, Sommers [13] proposed to parallel transport the second direction along the first principal geodesic to define the second coordinates, and iteratively define higher order coordinates through horizontal development along the previous modes.

All the cited methods are intrinsically forward methods that build successively larger and larger approximation spaces. A notable exception is the concept of Principal Nested Spheres (PNS), proposed by Jung, et al. [4] as a general framework for non-geodesic decomposition of high-dimensional spheres or high-dimensional planar landmarks shape spaces. Here, subsphere can be viewed as a slicing of a higher dimensional sphere by an affine hyperplane. In this process, the nested subsphere is not of radius one, unless the hyperplane is passing through the origin. The backward analysis approach, determining a decreasing family of subspace, has been recently generalized to more general manifold with the help of a nested sequence of relations [1]. However, up to know, such a sequence of relationships was only known for spheres or Euclidean spaces.

In [10], we proposed to replace geodesic subspaces with a new and more general type of family of subspaces in manifolds: barycentric subspaces (BS). They are implicitly defined as the locus of points which are weighted means of $k+1$ reference points. Depending on the generalization of the mean that we use on manifolds, Fréchet mean, Karcher mean or exponential barycenter, we obtain the Fréchet / Karcher / Exponential barycentric subspaces (FBS / KBS / EBS). These definition were shown to be included into each other. Here we call affine span the metric completion of the largest barycentric subspace. In generic conditions, barycentric subspaces are stratified spaces that are locally submanifolds of dimension k. Their singular set of dimension $k-l$ corresponds to the case where l of the reference point belongs to the barycentric subspace defined by the $k-l$ other reference points. In non-generic conditions, points may coalesce along certain directions, defining non local jets instead of a regular $k+1$-tuple. Geodesic subspaces (in a restricted sense), which are defined by k tangent vectors at a point, correspond to the limit of the affine span when the k-tuple converges towards that jet.

In this paper, we derive the equations of barycentric subspaces in one of the simplest manifold: the sphere. We show that the affine span of $k+1$ different reference points on the n-dimensional sphere is the k-dimensional great subsphere that contains the reference points. In fact, any $k+1$-tuple of points of a great k-dimensional subsphere generates the same affine span, which is also a geodesic subspace. This coincidence of spaces is due to the very high symmetry of the sphere. For second order jets, we show that we obtain subspheres of different radii as used in PNS. Among the points of the affine span, determining which ones belong to the Karcher barycentric subspaces (KBS) turns out to be a surprisingly difficult problem. Practical experiments show that the index of the variance at critical points can be arbitrary, thus subdividing the affine span into many regions. As a result, the KBS covers in general only a small portion of the subsphere containing the reference points. This suggests that the affine span might be a much more interesting definition for subspace definition pur-

poses. Finally, we discuss the use to these barycentric subspaces to generalize PCA on manifolds, a procedure that we name Barycentric Subspace Analysis (BSA). Barycentric subspaces can be naturally nested, by defining an ordering of the reference points. Like for PGA, this allows the construction of a forward nested sequence of subspaces which contains the Fréchet mean. However, BSA also allows the construction of backward nested sequence which may not contain the mean.

2 Introduction

In this section, we recall the main notations and results from of [10]. We consider an embedding Riemannian manifold $\mathcal{M}$ of dimension n. The Riemannian metric is denoted $\langle \, . \mid . \, \rangle_x$ on each tangent space $T_x\mathcal{M}$ of the manifold. The expression of the the underlying norm in a chart is $\|v\|_x^2 = v^{\mathrm{T}} \, G(x) \, v = v^i v^j g_{ij}(x)$ using Einstein notations for tensor contractions. We assume the manifold to be geodesically complete (no boundary nor any singular point that we can reach in a finite time). As an important consequence, the Hopf-Rinow-De Rham theorem states that there always exists at least one minimizing geodesic between any two points of the manifold.

We denote by $\exp_x(v)$ the *exponential map* at point x which associate to each tangent vector $v \in T_x\mathcal{M}$ the point of $\mathcal{M}$ reached by the geodesic starting at x with this tangent vector after a unit time. This map is a local diffeomorphism from $0 \in T_x\mathcal{M}$ to $\mathcal{M}$, and we denote $\overrightarrow{xy} = \log_x(y)$ its inverse: it may be defined as the smallest vector of $T_x\mathcal{M}$ that allows to shoot a geodesic from x to y. When the tangent space is provided with an orthonormal basis, the Riemannian exponential and logarithmic maps provide *a normal coordinate systems at* x. A set of normal coordinate systems at each point of the manifold realize an atlas which allows to work very easily on the manifold. The implementation of exp and log maps is the basis of programming on Riemannian manifolds, and most the geometric operations needed for statistics or image processing can be rephrased based on them [9, 11].

2.1 $(k+1)$-pointed Riemannian manifold

Let $\{x_0, \dots x_k\} \in \mathcal{M}^{k+1}$ be a set of $k+1$ distinct points in the Riemannian manifold $\mathcal{M}$ and $C(x_0, \dots x_k) = \cup_{i=0}^{k} C(x_i)$ be the union of the cut loci of these points. We call $(k+1)$-pointed manifold $\mathcal{M}^*(x_0, \dots x_k) = \mathcal{M} / C(x_0, \dots x_k)$ the submanifold of the non-cut points of the points.

Since the cut locus of each point is closed and has null measure, $\mathcal{M}^*(x_0, \dots x_k)$ is open and dense in $\mathcal{M}$. Thus, it is a submanifold of $\mathcal{M}$ (not necessarily connected). On this submanifold $\mathcal{M}^*(x_0, \dots x_k)$, the distance to the points x_i and the Riemannian log function $\overrightarrow{xx_i} = \log_x(x_i)$ are smooth.

2.2 Weighted moments of a $(k+1)$-pointed manifold

Let $(\lambda_0, \dots \lambda_k) \in \mathbb{R}^{k+1}$ be weights such that $\sum_i \lambda_i \neq 0$. We call such weights barycentric coordinates. They are elements of projective space $\mathcal{P}_k$ minus the orthogonal of the line element $\mathbb{1} = (1 : 1 : \dots 1)$: $\mathcal{P}_k^* = \{(\lambda_0 : \dots : \lambda_k) \in \mathbb{R}^{k+1}$ such that $\sum_i \lambda_i \neq 0\}$. Standard charts of this space are given either by the intersection of the line elements with the "upper" unit sphere S_k of $\mathbb{R}^{k+1}$ with north pole $\mathbb{1}/\sqrt{k}$ (unit weights) or by the k-plane of $\mathbb{R}^{k+1}$ passing through the point $\mathbb{1}/k$ and orthogonal to this vector. We call normalized weights $\underline{\lambda}_i = \lambda_i / (\sum_{j=0}^k \lambda_j)$ this last projection.

Given barycentric coordinates $\lambda \in \mathcal{P}_k^*$, we can consider the distribution (or 0-current) $\mu(x) = \sum_i \lambda_i \delta_{x_i}(x)$ on $\mathcal{M}$. As it is not normalized and weights can be negative, it is generally not a probability distribution. To define the the weighted n-order moment of that distribution, we have to restrict to the $(k+1)$-pointed Riemannian manifold $\mathcal{M}^*(x_0, \dots x_k)$ because the Riemannian log and distance functions are not defined (resp. smooth) at the cut-locus of the points $\{x_i\}$:

$$\mathfrak{M}_n(x, \lambda) = \sum_i \lambda_i \underbrace{\overrightarrow{xx_i} \otimes \overrightarrow{xx_i} \dots \otimes \overrightarrow{xx_i}}_{n \text{ times}} \tag{1}$$

The 0-th order moment (the mass) $\mathfrak{M}_0(\lambda) = \sum_i \lambda_i = \mathbb{1}^{\mathrm{T}} \lambda$ is constant. All other moment are homogeneous of degree 1 in λ and can be normalized by dividing by the mass $\mathfrak{M}_0(\lambda)$. The first order moment $\mathfrak{M}_1(x, \lambda) = \sum_i \lambda_i \overrightarrow{xx_i}$ is a smooth vector field on the manifold $\mathcal{M}^*(x_0, \dots x_k)$. The second and higher order moments are smooth $(n, 0)$ tensor fields that will be used through their contraction with the Riemannian curvature tensor.

2.3 Barycentric subspaces of $k+1$ points

Let $(\mathcal{M}, \mathrm{dist})$ be a metric space and $(x_0, \dots x_k) \in \mathcal{M}^k$ be $k+1$ distinct reference points. The (normalized) weighted variance at point x with weight $\lambda \in \mathcal{P}_k^*$ is: $\sigma^2(x, \lambda) = \frac{1}{2} \sum_{i=0}^k \underline{\lambda}_i \, \mathrm{dist}^2(x, x_i) = \frac{1}{2} \sum_{i=0}^k \lambda_i \, \mathrm{dist}^2(x, x_i) / (\sum_{j=0}^k \lambda_j)$. The **Fréchet barycentric subspace** is the locus of weighted Fréchet means of these points, i.e. the set of absolute minima of the weighted variance:

$$FBS(x_0, \dots x_k) = \{\arg\min_{x \in \mathcal{M}} \sigma^2(x, \lambda), \ \lambda \in \mathcal{P}_k^*\}.$$

The **Karcher barycentric subspace** $KBS(x_0, \dots x_k)$ is defined similarly with local minima instead of global ones.

A point $x \in \mathcal{M}^*(x_0, \dots x_k)$ is a weighted exponential barycenters of the reference points if we can find barycentric coordinates $\lambda \in \mathcal{P}_k^*$ such that

$$\mathfrak{M}_1(x, \lambda) = \sum_{i=0}^k \lambda_i \overrightarrow{xx_i} = 0. \tag{2}$$

The **Exponential barycentric subspace** $EBS(x_0, \dots x_k)$ is the set of weighted exponential barycenters of the reference points in $\mathcal{M}^*(x_0, \dots x_k)$:

$$EBS(x_0, \dots x_k) = \{x \in \mathcal{M}^*(x_0, \dots x_k) | \exists \lambda \in \mathcal{P}_k^* : \mathfrak{M}_1(x, \lambda) = 0\}.$$

This definition is only valid on $\mathcal{M}^*(x_0, \dots x_k)$ and may hide some discontinuities or continuity on the union of the cut locus of the reference points. In order to ensure the completeness of the subspace and potentially reconnect components, we define consider the closure of this set: we call **affine span** of the points $(x_0, \dots x_k) \in \mathcal{M}^k$ the closure of the EBS in $\mathcal{M}$:

$$\mathrm{Aff}(x_0, \dots x_k) = \overline{\mathrm{EBS}}(x_0, \dots x_k).$$

Because we assumed that $\mathcal{M}$ is geodesically complete, this is equivalent to the metric completion of the EBS.

Outside the cut locus of the reference points, which is of null measure, the gradient of the squared distance $d^2_{x_i}(x) = \mathrm{dist}^2(x, x_i)$ is well defined and is equal to $\nabla d^2_{x_i}(x) = -2 \log_x(x_i)$. Thus, one recognizes that Eq.(2) defines nothing else than the critical points of the variance $\sigma^2(x, \lambda) = \frac{1}{2} \sum_i \lambda_i \, \mathrm{dist}^2(x, x_i)$. The local minima of the variance which are potnetially located on the cut-locus of the reference points are not part of the EBS but they are recovered in the affine span thanks to the metric completion. FBS and KBS are thus included in the affine span, and the affine span is the largest of the barycentric subspaces.

2.4 SVD Characterization of the EBS

Let us consider field of $n \times (k+1)$ matrices $Z(x) = [\overrightarrow{xx_0}, \dots \overrightarrow{xx_k}]$ on $\mathcal{M}^*(x_0, \dots x_k)$. We can rewrite Eq.(2) in matrix form: $\mathfrak{M}_1(x, \lambda) = Z(x)\lambda = 0$. Thus, we see that the EBS is controled by the kernel of the matrix field $Z(x)$. Let now $Z(x) = U(x).S(x).V(x)^{\mathrm{T}}$ be a singular decomposition with singular values sorted in decreasing order. The barycentric subspace $EBS(x_0, \dots x_k)$ is the zero level-set of the $k+1$ singular value $s_{k+1}(x)$ and the subspace of valid barycentric weights is spanned by the right singular vectors corresponding to the l vanishing singular values: $Span(v_{k-l}, \dots v_k)$ (it is empty if $l = 0$).

2.5 Karcher barycentric subspace and positive span

A critical point of the variance $x \in EBS(x_0, \dots x_k)$ is said non-degenerated (resp. positive) if the Hessian matrix $H(x, \lambda) = -\sum_{i=0}^{k} \lambda_i D_x \log_x(x_i)$ is invertible (resp. positive definite) for all λ in the right singular space of the zero singular value of $Z(x)$. The set of degenerate (resp. non-degenerate or positive) exponential barycenter is called the degenerate EBS and denoted $EBS^0(x_0, \dots, x_k)$ (resp. non-degenerate $EBS^*(x_0, \dots, x_k)$ or positive $EBS^+(x_0, \dots x_k)$). In Euclidean spaces, all the points are positive and non-degenerated. However, in Riemannian manifolds, we generally have degenerated points and non-degenerated points which are non-positive, as we will see with with the example of spheres. Thus, we can conclude that the KBS is the positive EBS plus potentially some degenerate points of the affine span and some points of the cut locus of the reference points.

2.6 Geodesic subspaces as limit case of the affine span

The usual definition of the geodesic subspaces $GS(W_x) = \{\exp_x(w), w \in W_x\}$ is too large in certain cases to be useful (e.g. in torus). We call restricted geodesic submanifold $GS^*(W_x) = \{\exp_x(w), w \in W_x \cap D(x)\}$ its restriction to the points that are reached without going through the cut locus of x. This is a well defined submanifold of $\mathcal{M}$ whose points are described by homogeneous coordinates at infinity (or on the equator of $\mathbb{1}/\sqrt{k}$ depending of the chart we chose for $\mathcal{P}_k^*$) of the affine span $\mathrm{Aff}(x, x_1, \dots x_k)$ when the points $x_i = \exp_x(\eta w_i)$ are converging to x at first order along the tangent vectors w_i defining the k-dimensional subspace $W_x \subset T_x\mathcal{M}$.

3 Example on spheres

Let us consider the unit sphere as our base manifold. We represent points of $\mathcal{M} = \mathcal{S}_n$ as unit vectors in $\mathbb{R}^{n+1}$. The tangent space at x is naturally represented by the linear space of vectors orthogonal to x: $T_x\mathcal{S}_n = \{v \in \mathbb{R}^{n+1}, v^{\mathrm{T}}x = 0\}$. The natural Riemannian metric on the unit sphere is inherited from the Euclidean metric of the embedding space $\mathbb{R}^{n+1}$. With these conventions, the Riemannian distance is the arc-length $d(x, y) = \arccos(x^{\mathrm{T}}y)$. Let us denote $f(\theta) = 1/\mathrm{sinc}(\theta) = \theta/\sin(\theta)$. The spherical exp and log maps are

$$\exp_x(v) = \cos(\|v\|)x + \frac{\sin(\|v\|)}{\|v\|}v = \cos(\|v\|)x + \mathrm{sinc}(\|v\|)v$$

$$\log_x(y) = \frac{\theta}{\sin(\theta)}(y - \cos(\theta)x) = f(\theta)(y - \cos(\theta)x) \quad \text{with} \quad \theta = \arccos(x^{\mathrm{T}}y).$$

Notice that $f(\theta)$ is a smooth function from $]-\pi;\pi[$ to $\mathbb{R}$ that is always greater than one and is locally quadratic at zero: $f(\theta) = 1 + \theta^2/6 + O(\theta^4)$.

Let us pick $k+1$ points on the sphere that we put in a matrix $X = [x_0; \dots x_k]$. In the sequel, we use the same notation for the matrix and the set of points. The cut locus of x_i is its antipodal point $-x_i$ so that $\mathcal{M}^*(X) = \mathcal{S}_n/-X$. Let us denote $\theta_i = \arccos(x_i^{\mathrm{T}}x)$. The log at a point x is $\overrightarrow{xx_i} = (\mathrm{Id} - xx^{\mathrm{T}})f(\theta_i)x_i$, so that the first weighted moment is

$$\mathfrak{M}_1(x, \lambda) = (\mathrm{Id} - xx^{\mathrm{T}})\sum_i \lambda_i f(\theta_i)x_i = (\mathrm{Id} - xx^{\mathrm{T}})XF(X, x)\lambda$$

where $F(X, x) = \mathrm{Diag}(f(\theta_i))$ is a diagonal matrix with entries that are always greater than one for $x \in \mathcal{M}^*(X)$. Thus the matrix $Z(x) = (\mathrm{Id} - xx^{\mathrm{T}})XF(X, x)$ has the same kernel as the matrix $\tilde{Z}(x) = (\mathrm{Id} - xx^{\mathrm{T}})X$. This corresponds to a renormalized $\tilde{\lambda} = F(X, x)\lambda$ of the weights which is linear in λ but non-linear in x and X through the function $F(X, x)$. The solutions of the equation $\tilde{Z}(x)\tilde{\lambda} = 0$ under the constraint $\|x\| = 1$ are given by $(x^{\mathrm{T}}X\tilde{\lambda})x = X\tilde{\lambda}$ or more explicitly $x = \pm X\tilde{\lambda}/\|X\tilde{\lambda}\|$. This means that the point $x \in \mathcal{M}^*(X)$ has to belong to the Euclidean span of the reference vectors. Conversely, any unit vector $x = X\tilde{\lambda}$ of

the Euclidean span of X verifies $\tilde{Z}(x)\tilde{\lambda} = 0$. Thus, the unit vectors $x = \pm X\tilde{\lambda}/\|\tilde{\lambda}\|$ have barycentric coordinates $\lambda = F(X,x)^{(\text{-}1)}\tilde{\lambda}$ if x is not at the cut locus of the reference points. This shows that

$$\text{Aff}(X) = \text{Span}\{x_0, \dots x_k\} \cap \mathcal{S}_n. \tag{3}$$

Following the same principle, we can orthogonalize the reference points x_i: let us denote by $X = USV^{\mathrm{T}}$ a singular value decomposition of the matrix of reference vectors. All the singular values s_i are positive since the vectors x_i are assumed to be linearly independent. Thus, $Z(x)$ has the same kernel as $Y(x) = (\text{ Id} - xx^{\mathrm{T}})U$. This shows that the exponential barycentric subspace generated by the original and orthogonalized points is the same, except at the cut locus of all these points which however belongs to the closure: the affine span. Thus, for spherical data as for Euclidean data, the affine span only depend on the reference points through the point of the Grassmanian that they define.

When the reference points x_i are not linearly independent, the matrix X has $l \geq 1$ vanishing singular values. A singular value decomposition $X = USV^{\mathrm{T}}$ shows that the value of $\tilde{\lambda}$ (and thus of λ) is in that case unconstrained in the vector space generated by the right singular vectors $v_{k-l}, \dots v_k$ associated to the l vanishing singular values $s_{k-l}, \dots s_k$. Thus, the space of admissible weights at each point of the EBS is of dimension l, and the affine span itself is still the subsphere generated the Euclidean span of the reference points (minus their cut-locus), which is of dimension $k - l$.

In conclusion, the affine span $\text{Aff}(X)$ of $k+1$ different reference unit points $X = [x_0; \dots x_k]$ on the n-dimensional sphere $\mathcal{S}_n$ provided with the canonical Euclidean metric of the embedding space $\mathbb{R}^{n+1}$ is the great subsphere of dimension $rank(X) - 1$ that contains the reference points.

3.1 Reference points coalescing at order 1

Assume now all the reference points coalesce to a single point $x_i = \exp_{x_0}(\epsilon w_i)$ along the tangent vectors w_i which are satisfying $x_0^{\mathrm{T}} w_i = 0$ (to belong to the tangent space at x_0) and $\sum_i w_i = 0$. This amounts to say that we are following the curve $X_\epsilon = X_0 + \epsilon W$ in the space of barycentric subspaces, with $X_0^{\mathrm{T}} W = 0$ and $W X_0^{\mathrm{T}} = 0$, where here $X_0 = x_0 \mathbb{1}^{\mathrm{T}}$.

As previously, the points of the EBS of X_ϵ are solution of the equation $Z(x)\lambda = 0$ and since $Z(x)$ has the same kernel as the matrix $Z'(x) = (\text{ Id} - xx^{\mathrm{T}})X_\epsilon$, x is solution if we can find some α_ϵ and λ such that $\alpha_\epsilon x = X_\epsilon \lambda$. The additional constraint $\|x\|^2 = 1$ tells us that $\alpha_\epsilon^2 = \|X_\epsilon \lambda\|^2$, which gives $x = X_\epsilon \lambda / \alpha_\epsilon$ when we take the positive root and reintegrate the sign into λ.

In our case, thanks to the orthogonality of X_0 and W, we have

$$\alpha_\epsilon = \|X_0\lambda + \epsilon W\lambda\| = \sqrt{\|X_0\lambda\|^2 + \epsilon^2 \|W\lambda\|^2},$$

Assuming that $\|X_0\lambda\| \neq 0$ (which is in particular the case when $X_0 = x_0\mathbb{1}^{\mathrm{T}}$ since $\mathbb{1}^{\mathrm{T}}\lambda$ is assumed not to vanish), we have $\alpha_\epsilon^{-1} = \frac{1}{\|X_0\lambda\|} - \frac{\epsilon^2}{2}\frac{\|W\lambda\|^2}{\|X_0\lambda\|^3} + O(\epsilon^4)$,

so that finally

$$x = \left(1 - \frac{\epsilon^2}{2}\frac{\|W\lambda\|^2}{\|X_0\lambda\|^2} + O(\epsilon^4)\right)\left(\frac{X_0\lambda}{\|X_0\lambda\|} + \epsilon\frac{W\lambda}{\|X_0\lambda\|}\right) = \frac{X_0\lambda}{\|X_0\lambda\|} + W\frac{\epsilon\lambda}{\|X_0\lambda\|} + O(\epsilon^2)$$

Thus, we see that the space $Aff(x)$ is the intersection of the sphere with the Euclidean hyperplane generated by X_0 and W, which is once again the geodesic subspace $GS(X_0, W)$.

3.2 Coalescence at order 2 and link with principal nested spheres

Principal nested spheres were proposed by Jung, Dryden and Marron as a general framework for non-geodesic decomposition of high-dimensional spheres or high-dimensional planar landmarks shape spaces [5, 4]. A subsphere $\mathcal{A}_{n-1}$ of $\mathcal{S}_n$ is defined as the set of points which are at a fixed distance $\theta \in (0, \pi/2]$ of a point $x \in \mathcal{S}_n$: $\mathcal{A}_{n-1}(x, \theta) = \{y \in \mathcal{S}_n \,/\, d(x, y) = \theta\}$. The subsphere $\mathcal{A}_{n-1}(x, \theta)$ can be viewed as a slicing of $\mathcal{S}_n$ by the n-dimensional affine hyperplane $P(x, \theta) = \{y \in \mathbb{R}^{n+1} \,/\, y^{\mathrm{T}}x = \cos(\theta)\}$. Notice that the coordinates $(x, \cos(\theta))$ of the affine hyperplane parametrize all the possible subspheres of dimension $n - 1$. In this process, the nested subsphere is not of radius one, unless one takes $\theta = \pi/2$, in which case the hyperplane is passing through the origin.

In order to figure out how nested subspheres and barycentric subspaces are related, we consider the top circle of radius $r \in [0; 1]$ around the axis e_3 on the 3-sphere implicitly described by the equation $x^{\mathrm{T}}e_3 = \sqrt{1 - r^2}$. The computations would be exactly the same for the bottom circle $x^{\mathrm{T}}e_3 = -\sqrt{1 - r^2}$. The explicit equation of our circle is: $x(\psi) = r\cos(\psi)e_1 + r\sin(\psi)e_2 + \sqrt{1 - r^2}e_3$. We consider the three points on this cicle at angle $\psi = 0$, $\psi = \epsilon$ and $\psi = -\epsilon$: $X = [x(0), x(\epsilon), x(-\epsilon)]$. The spherical affine span consists of the points $x = X\lambda/\|X\lambda\|$ for $\lambda \in \mathcal{S}_3$, with:

$$\begin{aligned} X\lambda &= r\left(\lambda_0 + \cos(\epsilon)(\lambda_1 + \lambda_2)\right)e_1 + r(\lambda_1 - \lambda_2)\sin(\epsilon)e_2 + \sqrt{1 - r^2}(\lambda_0 + \lambda_1 + \lambda_2)e_3 \\ &= (\lambda_0 + \lambda_1 + \lambda_2)\left(re_1 + \sqrt{1 - r^2}e_3\right) + \epsilon r(\lambda_1 - \lambda_2)e_2 - \frac{r\epsilon^2}{2}(\lambda_1 + \lambda_2)e_1 + O(\epsilon^3). \end{aligned}$$

Using new coordinates $s = (\lambda_0 + \lambda_1 + \lambda_2)$, $u = (\lambda_1 - \lambda_2)\epsilon/s$ and $v = (\lambda_1 + \lambda_2)\epsilon^2/(2s)$, we get the equation

$$X\frac{\lambda}{s} = r(1 - v)e_1 + rue_2 + \sqrt{1 - r^2}e_3 + O(\epsilon^3).$$

Thus, the equation $x = X\lambda$ can only describe the hyperplane $x^{\mathrm{T}}e_3 = \sqrt{1 - r^2}$ when ϵ goes to zero (up to a scaling factor s that we can freely choose to be 1 thanks to the homogeneous coordinates), and its intersection with the sphere can only describe a circle of radius r.

Iterating the process, one can generalize the above construction to subspheres of arbitrary dimensions. Thus, we see that Nested Spheres as a limit case of the affine span when the k reference points tend to a 2-jet. It would be interesting to determine which types of subspaces could be obtained by such limits for more general non-local and higher order jets.

4 Barycentric subspace analysis

We turn in this section to the generalization of principal component analysis itself. PCA can be viewed as the search for a sequence of nested linear spaces that best approximate the data at each level, for instance by minimizing the variance of the residues. In a Euclidean space, this process boils down to an independent optimization of orthogonal subspaces at each level of approximation, thanks to the Pythagorean theorem. This allows to build each subspace of the sequence by adding (resp. subtracting) the optimal one-dimensional subspace iteratively in a forward (resp. backward) analysis. Of course, this property does not scale up to manifolds, for which subspaces have no reason to be orthogonal (even this notion is not well defined).

4.1 Flags of barycentric subspaces in manifolds

Nestedness of approximation spaces has been argued to be one of the most important characteristics for generalizing PCA to more general spaces [1]. Barycentric submanifolds can easily be nested, for instance by adding or removing one or several points at a time, which corresponds to put the barycentric weight of this (or these) point(s) to zero. One obtains in that case a family of embedded submanifolds which we call a flag because this generalizes flags of vector spaces. Indeed a flag of a vector space V is a filtration of subspaces (an increasing sequence of subspaces, where "increasing" means each subspace is a proper subspace of the next): $\{0\} = V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_k = V$. Denoting by $d_i = \dim(V_i)$ the dimension of the subspaces, we have $0 = d_0 < d_1 < d_2 < \cdots < d_k = n$, where n is the dimension of V. Hence, we must have $k \leq n$. A flag is called a complete flag if $d_i = i$, otherwise it is called a partial flag.

With barycentric subspaces of an n-dimensional manifold $\mathcal{M}$, an ordering of $n+1$ distinct points $x_0, \dots x_n$ defines a complete flag of barycentric subspaces in the sense that: $BS(x_0) = \{x_0\} \subset \cdots BS(x_0, x_1, x_k) \cdots \subset BS(x_0, \dots x_n)$.

Grouping points together in the addition/removal process generates a partial flag of barycentric subspaces. Among the barycentric subspaces, the affine span seems to be the most interesting definition to use because the affine span of $n+1$ distinct points covers the full manifold: $\mathrm{Aff}(x_0, \dots x_n) = \mathcal{M}$ while we only generate a submanifold with the Fréchet or Karcher barycentric subspaces, as we have seen with the example of spheres.

4.2 Forward and backward barycentric subspaces analysis

In the classical PCA, the flag of linear subspaces can be built in a forward way, by computing the best 0-th order approximation (the mean), then the best first order approximation (the first mode), etc. It can also be built backward, by removing the direction with the minimal residue from the current affine subspace. In a manifold, we can use similar forward and backward analysis, but they have no reason to give the same result.

With a forward analysis, we compute iteratively the flag of affine spans by adding one point at a time and keeping the previous ones fixed. Thus, we begin by computing the optimal barycentric subspace of dimension 0: $\mathrm{Aff}(x_0) = \{x_0\}$. Since there is only one weight and it should be unit, the optimal point x_0 found by minimizing the unexplained variance is a Karcher mean. Adding one more point amounts to compute the geodesic passing through the mean that best approximate the data. Adding a second points now differ from PGA, unless the three points coalesce to a single one. The procedure is continued point by point, which mean that the Fréchet mean always belong to the barycentric subspace. In practice, the forward analysis should be stopped when the variance of the residues reaches the noise level of the data, hopefully with k much lower than the dimension n of the embedding manifold, which allows to have an efficient dimension reduction.

The backward analysis consists in iteratively removing one dimension, thus one point in our case. One theoretically should start with a full set of points $x_0, \dots x_n$ which generates the full manifold and chose which one to remove. However, as all the sets of $n+1$ distinct points generate the full manifold, the optimization really begin with the set of n points $x_0, \dots x_{n-1}$. Actually this should normally be the only time when we perform an optimization for the point positions, since one should afterward only test for which of the n points we should remove, and this optimization is particularly ill-posed and inefficient in very large dimensional spaces! In order to get around this problem, we propose to run a non-nested forward analysis until we reach the noise level of the data for a dimension $k \gg n$. Since the goal is only to characterize the optimal k-dimensional subspace, we may optimize the point positions at each step to better fit the data. Then, a backward sweep at the end only reorders the points if necessary by iteratively selecting the one that least increase the unexplained variance. With this process, there is no reason why the Fréchet mean should belong to the reference points (and even to any of the barycentric subspaces). For instance, if we have clusters of points, one expects the reference points to localize within these clusters rather than at the Fréchet mean.

5 Discussion

We have first investigated in the paper barycentric subspaces in spheres and shown that they encompass both principal geodesic subspaces and nested subspheres as limit cases. It would be interesting to see if we can obtain other types of subspaces with higher order and non-local jets.

The second study point of this paper concerns the generalization of PCA to manifolds using Barycentric Subspace Analysis (BSA). We showed that an ordering of the reference points naturally defines a flag of nested barycentric subspaces. We proposed a first optimization procedure, but the lack of symmetry between the forward and the backward estimations calls for a proper global criterion to be optimized by all k-tuple for $k = 0 \dots n$ together and not just a greedy approach as done by the classical forward and backward approaches.

Other potential practical issues include the fact that the optimization on k-tuple might have multiple solutions, as in the case of spheres. Here, we need to find a suitable quotient space similar to the quotient definition of Grassmanians. The optimization might also converge towards a non-local jet instead on a k-tuple, and good renormalization techniques need to be designed to guaranty the numerical stability.

References

1. J. Damon and J. S. Marron. Backwards Principal Component Analysis and Principal Nested Relations. *Journal of Mathematical Imaging and Vision*, 50(1-2):107–114, Oct. 2013.
2. P. Fletcher, C. Lu, S. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, Aug. 2004.
3. M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, 10:215–310, 1948.
4. S. Jung, I. L. Dryden, and J. S. Marron. Analysis of principal nested spheres. *Biometrika*, 99(3):551–568, Sept. 2012.
5. S. Jung, X. Liu, J. S. Marron, and S. M. Pizer. Generalized PCA via the Backward Stepwise Approach in Image Analysis. In J. Angeles, B. Boulet, J. J. Clark, J. Kvecses, and K. Siddiqi, editors, *Brain, Body and Machine*, number 83 in Advances in Intelligent and Soft Computing, pages 111–123. Springer Berlin Heidelberg, 2010.
6. H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications in Pure and Applied Mathematics*, 30:509–541, 1977.
7. H. Karcher. Riemannian Center of Mass and so called karcher mean. *arXiv:1407.2087 [math]*, July 2014. arXiv: 1407.2087.
8. W. Kendall. Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. *Proc. London Math. Soc.*, 61(2):371–406, 1990.
9. X. Pennec. Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006. A preliminary appeared as INRIA RR-5093, January 2004.
10. X. Pennec. Barycentric Subspaces and Affine Spans in Manifolds. In *Geometric Science of Information GSI'2015*, Proceedings of Geometric Science of Information GSI'2015, Palaiseau, France, Oct. 2015.
11. X. Pennec, P. Fillard, and N. Ayache. A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, 66(1):41–66, 2006. A preliminary version appeared as INRIA Research Report 5255, July 2004.
12. A. M. S. Huckemann, T. Hotz. Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions. *Statistica Sinica*, pages 1–100, 2010.
13. S. Sommer. Horizontal Dimensionality Reduction and Iterated Frame Bundle Development. In F. Nielsen and F. Barbaresco, editors, *Geometric Science of Information*, number 8085 in Lecture Notes in Computer Science, pages 76–83. Springer Berlin Heidelberg, 2013.
14. S. Sommer, F. Lauze, and M. Nielsen. Optimization over geodesics for exact principal geodesic analysis. *Advances in Computational Mathematics*, 40(2):283–313, June 2013.