

BARYCENTRIC SUBSPACE ANALYSIS ON MANIFOLDS

BY XAVIER PENNEC¹

Université Côte d'Azur and INRIA

This paper investigates the generalization of Principal Component Analysis (PCA) to Riemannian manifolds. We first propose a new and general type of family of subspaces in manifolds that we call barycentric subspaces. They are implicitly defined as the locus of points which are weighted means of $k + 1$ reference points. As this definition relies on points and not on tangent vectors, it can also be extended to geodesic spaces which are not Riemannian. For instance, in stratified spaces, it naturally allows principal subspaces that span several strata, which is impossible in previous generalizations of PCA. We show that barycentric subspaces locally define a submanifold of dimension k which generalizes geodesic subspaces.

Second, we rephrase PCA in Euclidean spaces as an optimization on flags of linear subspaces (a hierarchy of properly embedded linear subspaces of increasing dimension). We show that the Euclidean PCA minimizes the Accumulated Unexplained Variances by all the subspaces of the flag (AUV). Barycentric subspaces are naturally nested, allowing the construction of hierarchically nested subspaces. Optimizing the AUV criterion to optimally approximate data points with flags of affine spans in Riemannian manifolds lead to a particularly appealing generalization of PCA on manifolds called Barycentric Subspace Analysis (BSA).

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 double interpretation is available through Pythagoras' theorem, which does not hold in more general manifolds. A second important observation is that principal components of different orders are nested, enabling the forward or backward construction of nested principal components.

Generalizing PCA to manifolds first requires the definition of the equivalent of affine subspaces in manifolds. For the zero-dimensional subspace, an 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 (1948) in general

Received June 2016; revised July 2017.

¹Supported in part by the Erwin Schrödinger Institute in Vienna through a three-weeks stay in February 2015 during the program Infinite-Dimensional Riemannian Geometry with Applications to Image Matching and Shape and by the Inria Associated team GeomStats between Asclepios and Holmes' lab at the Stanford Statistics Department.

MSC2010 subject classifications. Primary 60D05, 62H25; secondary 58C06, 62H11.

Key words and phrases. Manifold, Fréchet mean, barycenter, flag of subspaces, PCA.

metric spaces. For simply connected Riemannian manifolds of nonpositive curvature, the minimum is unique and is called the Riemannian center of mass. This fact was already known by Cartan in the 1920s, but was not used for statistical purposes. [Karcher \(1977\)](#) and [Buser and Karcher \(1981\)](#) first established conditions on the support of the distribution to ensure the uniqueness of a local minimum in general Riemannian manifolds. This is now generally called Karcher mean, although there is a dispute on the naming [[Karcher \(2014\)](#)]. From a statistical point of view, [Bhattacharya and Patrangenaru \(2003, 2005\)](#) have studied in depth the asymptotic properties of the empirical Fréchet/Karcher means.

The one-dimensional component can naturally be a geodesic passing through the mean point. Higher-order components are more difficult to define. The simplest generalization is tangent PCA (tPCA), which amounts unfolding the whole distribution in the tangent space at the mean, and computing the principal components of the covariance matrix in the tangent space. The method is thus based on the maximization of the explained variance, which is consistent with the entropy maximization definition of a Gaussian on a manifold proposed by [Pennec \(2006\)](#). tPCA is actually implicitly used in most statistical works on shape spaces and Riemannian manifolds because of its simplicity and efficiency. 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 distributions which are multimodal or supported on large compact subspaces (e.g., circles or spheres).

Instead of an analysis of the covariance matrix, [Fletcher et al. \(2004\)](#) proposed the minimization of squared distances to subspaces which are totally geodesic at a point, a procedure coined Principal Geodesic Analysis (PGA). These Geodesic Subspaces (GS) are spanned by the geodesics going through a point with tangent vector restricted to a linear subspace of the tangent space. However, the least-squares procedure is computationally expensive, so that the authors approximated it in practice with tPCA, which led to confusions between tPCA and PGA. A real implementation of the original PGA procedure was only recently provided by [Sommer, Lauze and Nielsen \(2014\)](#). PGA is allowing to build a flag (sequences of embedded subspaces) of principal geodesic subspaces consistent with a forward component analysis approach. Components are built iteratively from the mean point by selecting the tangent direction that optimally reduces the square distance of data points to the geodesic subspace. In this procedure, the mean always belongs to geodesic subspaces even when it is outside of the distribution support.

To alleviate this problem, [Huckemann and Ziezold \(2006\)](#), and later [Huckemann, Hotz and Munk \(2010\)](#), proposed to start at the first-order component directly with the geodesic best fitting 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, [Sommer \(2013\)](#) proposed a parallel transport of 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.

These are all intrinsically forward methods that build successively larger approximation spaces for the data. A notable exception is the concept of Principal Nested Spheres (PNS), proposed by Jung, Dryden and Marron (2012) in the context of planar landmarks shape spaces. A backward analysis approach determines a decreasing family of nested subspheres by slicing a higher dimensional sphere with affine hyperplanes. In this process, the nested subspheres are not of radius one, unless the hyperplanes pass through the origin. Damon and Marron (2013) have recently generalized this approach to manifolds with the help of a “nested sequence of relations”. However, up to now, such a sequence was only known for spheres or Euclidean spaces.

We first propose in this paper new types of family of subspaces in manifolds: barycentric subspaces generalize geodesic subspaces and can naturally be nested, allowing the construction of inductive forward or backward nested subspaces. We then rephrase PCA in Euclidean spaces as an optimization on flags of linear subspaces (a hierarchy of properly embedded linear subspaces of increasing dimension). To that end, we propose an extension of the unexplained variance criterion that generalizes nicely to flags of barycentric subspaces in Riemannian manifolds. This leads to a particularly appealing generalization of PCA on manifolds: Barycentric Subspace Analysis (BSA).

Paper organization. We recall in Section 2 the notions and notation needed to define statistics on Riemannian manifolds, and we introduce the two running example manifolds of this paper: n -dimensional spheres and hyperbolic spaces. Exponential Barycentric Subspaces (EBS) are then defined in Section 3 as the locus of weighted exponential barycenters of $k + 1$ affinely independent reference points. The closure of the EBS in the original manifold is called affine span [this differs from the preliminary definition of Pennec (2015)]. Equations of the EBS and affine span are exemplified on our running examples: the affine span of $k + 1$ affinely independent reference points is the great subsphere (resp., sub-hyperbola) that contains the reference points. In fact, other tuple of points of that subspace generates the same affine span, which is also a geodesic subspace. This coincidence is due to the very high symmetry of the constant curvature spaces.

Section 4 defines the Karcher (resp., Fréchet) barycentric subspaces (KBS, resp., FBS) as the local (resp., global) minima of the weighted squared distance to the reference points. As the definitions relies on distances between points and not on tangent vectors, they are also valid in more general non-Riemannian geodesic spaces. For instance, in stratified spaces, barycentric subspaces may naturally span several strata. For Riemannian manifolds, we show that our three definitions are subsets of each other (except possibly at the cut locus of the reference points):

the largest one, the EBS, is composed of the critical points of the weighted variance. It forms a cell complex according to the index of the critical points. Cells of positive index gather local minima to form the KBS. We explicitly compute the Hessian on our running spherical and hyperbolic examples. Numerical tests show that the index can be arbitrary, thus subdividing the EBS into several regions for both positively and negatively curved spaces. Thus, the KBS consistently covers only a small portion of the affine span in general and is a less interesting definition for subspace analysis purposes.

For affinely independent points, we show in Section 5 that the regular part of a barycentric subspace is a stratified space which is locally a submanifold of dimension k . At the limit, points may coalesce along certain directions, defining nonlocal jets² instead of a regular $k + 1$ -tuple. Restricted geodesic subspaces, which are defined by k vectors tangent at a point, correspond to the limit of the affine span when the k -tuple converges towards that jet.

Finally, we discuss in Section 6 the use of these barycentric subspaces to generalize PCA on manifolds. Barycentric subspaces can be naturally nested by defining an ordering of the reference points. Like for PGA, this enables the construction of a forward nested sequence of subspaces which contains the Fréchet mean. In addition, BSA also provides backward nested sequences which may not contain the mean. However, the criterion on which these constructions are based can be optimized for each subspace independently but not consistently for the whole sequence of subspaces. In order to obtain a global criterion, we rephrase PCA in Euclidean spaces as an optimization on flags of linear subspaces (a hierarchies of properly embedded linear subspaces of increasing dimension). To that end, we propose an extension of the unexplained variance criterion [the Accumulated Unexplained Variance (AUV) criterion] that generalizes nicely to flags of affine spans in Riemannian manifolds. This results into a particularly appealing generalization of PCA on manifolds that we call Barycentric Subspace Analysis (BSA).

2. Riemannian geometry. In statistics, directional data occupy a place of choice [Dryden (2005), Huckemann and Ziezold (2006)]. Hyperbolic spaces are also the simplest models of negatively curved spaces which model the space of isotropic Gaussian parameters with the Fisher–Rao metric in information geometry [Costa, Santos and Strapasson (2015)]. As nonflat constant curvature spaces, both spherical and hyperbolic spaces are now considered in manifold learning for embedding data [Wilson et al. (2014)]. Thus, they are ideal examples to illustrate the theory throughout this paper.

² p -jets are equivalent classes of functions up to order p . Thus, a p -jet specifies the Taylor expansion of a smooth function up to order p . Nonlocal jets, or multijets, generalize subspaces of the tangent spaces to higher differential orders with multiple base points.

2.1. Tools for computing on Riemannian manifolds. We consider a differential manifold \mathcal{M} endowed with a smooth scalar product $\langle \cdot | \cdot \rangle_x$ called the Riemannian metric on each tangent space $T_x\mathcal{M}$ at point x of \mathcal{M} . In a chart, the metric is specified by the dot product of the tangent vector to the coordinate curves: $g_{ij}(x) = \langle \partial_i | \partial_j \rangle_x$. The Riemannian distance between two points is the infimum of the length of the curves joining these points. Geodesics, which are critical points of the energy functional, are parametrized by arc-length in addition to optimizing the length. We assume here that the manifold is geodesically complete, that is, the definition domain of all geodesics can be extended to \mathbb{R} . This means that the manifold has 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 (i.e., whose length is the distance between the two points).

Normal coordinate system. From the theory of second-order differential equations, we know that there exists one and only one geodesic $\gamma_{(x,v)}(t)$ starting from the point x with the tangent vector $v \in T_x\mathcal{M}$. The exponential map at point x maps each tangent vector $v \in T_x\mathcal{M}$ to the point of the manifold that is reached after a unit time by the geodesic: $\exp_x(v) = \gamma_{(x,v)}(1)$. The exponential map is locally one-to-one around 0: we denote by $\overrightarrow{xy} = \log_x(y)$ its inverse. The injectivity domain is the maximal domain $D(x) \subset T_x\mathcal{M}$ containing 0 where the exponential map is a diffeomorphism. This is a connected star-shape domain limited by the tangential cut locus $\partial D(x) = C(x) \subset T_x\mathcal{M}$ [the set of vectors tv where the geodesic $\gamma_{(x,v)}(t)$ ceases to be length minimizing]. The cut locus $\mathcal{C}(x) = \exp_x(C(x)) \subset \mathcal{M}$ is the closure of the set of points where several minimizing geodesics starting from x meet. The image of the domain $D(x)$ by the exponential map covers all the manifold except the cut locus, which has null measure. Provided with an orthonormal basis, \exp and \log maps realize a normal coordinate system at each point x . Such an atlas is the basis of programming on Riemannian manifolds as exemplified in [Pennec, Fillard and Ayache \(2006\)](#).

Hessian of the squared Riemannian distance. On $\mathcal{M} \setminus C(y)$, the Riemannian gradient $\nabla^a = g^{ab}\partial_b$ of the squared distance $d_y^2(x) = \text{dist}^2(x, y)$ with respect to the fixed point y is $\nabla d_y^2(x) = -2\log_x(y)$. The Hessian operator (or double covariant derivative) ∇^2 is the covariant derivative of the gradient. In a normal coordinate at the point x , the Christoffel symbols vanish at x so that the Hessian of the square distance can be expressed with the standard differential D_x with respect to the footpoint x : $\nabla^2 d_y^2(x) = -2(D_x \log_x(y))$. It can also be written in terms of the differentials of the exponential map as $\nabla^2 d_y^2(x) = (D \exp_x |_{\overrightarrow{xy}})^{-1} D_x \exp_x |_{\overrightarrow{xy}}$ to explicitly make the link with Jacobi fields. Following [Brewin \(2009\)](#), we computed in [Pennec \(2017a\)](#) the Taylor expansion of this matrix in a normal coordinate

system at x :

$$(1) \quad [D_x \log_x(y)]_b^a = -\delta_b^a + \frac{1}{3} R_{cbd}^a \vec{x}\vec{y}^c \vec{x}\vec{y}^d + \frac{1}{12} \nabla_c R_{dbe}^a \vec{x}\vec{y}^c \vec{x}\vec{y}^d \vec{x}\vec{y}^e + O(\varepsilon^3).$$

Here, $R_{cbd}^a(x)$ are the coefficients of the curvature tensor at x and Einstein summation convention implicitly sums upon each index that appear up and down in the formula. Since we are in a normal coordinate system, the zeroth-order term is the identity matrix, like in Euclidean spaces, and the first-order term vanishes. The Riemannian curvature tensor appears in the second-order term and its covariant derivative in the third-order term. Curvature is the leading term that makes this matrix departing from the identity (the Euclidean case) and may lead to the noninvertibility of the differential.

Moments of point distributions. Let $\{x_0, \dots, x_k\}$ be a set of $k + 1$ points on a Manifold provided with weights $(\lambda_0, \dots, \lambda_k)$ that do not sum to zero. We may see these weighted points as the weighted sum of Diracs $\mu(x) = \sum_i \lambda_i \delta_{x_i}(x)$. As this distribution is not normalized and weights can be negative, it is generally not a probability. It is also singular with respect to the Riemannian measure. Thus, we have to take care in defining its moments as the Riemannian log and distance functions are not smooth at the cut-locus.

DEFINITION 1 [$(k + 1)$ -pointed/punctured Riemannian manifold]. Let $\{x_0, \dots, x_k\} \in \mathcal{M}^{k+1}$ be a set of $k + 1$ reference points in the n -dimensional Riemannian manifold \mathcal{M} and $C(x_0, \dots, x_k) = \bigcup_{i=0}^k C(x_i)$ be the union of the cut loci of these points. We call the object consisting of the smooth manifold \mathcal{M} and the $k + 1$ reference points a $(k + 1)$ -pointed manifold. Likewise, we call the submanifold $\mathcal{M}^*(x_0, \dots, x_k) = \mathcal{M} \setminus C(x_0, \dots, x_k)$ of the noncut points of the $k + 1$ reference points a $(k + 1)$ -punctured manifold.

On $\mathcal{M}^*(x_0, \dots, x_k)$, the distance to the points $\{x_0, \dots, x_k\}$ is smooth. The Riemannian log function $\vec{x}\vec{x}_i = \log_x(x_i)$ is also well defined for all the points of $\mathcal{M}^*(x_0, \dots, x_k)$. Since the cut locus of each point is closed and has null measure, the punctured manifold $\mathcal{M}^*(x_0, \dots, x_k)$ is open and dense in \mathcal{M} , which means that it is a submanifold of \mathcal{M} . However, this submanifold is not necessarily connected. For instance in the flat torus $(S_1)^n$, the cut-locus of $k + 1 \leq n$ points divides the torus into k^n disconnected cells.

DEFINITION 2 [Weighted moments of a $(k + 1)$ -pointed manifold]. Let $(\lambda_0, \dots, \lambda_k) \in \mathbb{R}^{k+1}$ such that $\sum_i \lambda_i \neq 0$. We call $\underline{\lambda}_i = \lambda_i / (\sum_{j=0}^k \lambda_j)$ the normalized weights. The weighted p th-order moment of a $(k + 1)$ -pointed Riemannian manifold is the p -contravariant tensor

$$(2) \quad \mathfrak{M}_p(x, \lambda) = \sum_i \lambda_i \underbrace{\vec{x}\vec{x}_i \otimes \vec{x}\vec{x}_i \otimes \dots \otimes \vec{x}\vec{x}_i}_{p \text{ times}}.$$

The normalized p th-order moment is $\underline{\mathfrak{M}}_p(x, \lambda) = \mathfrak{M}_p(x, \underline{\lambda}) = \mathfrak{M}_p(x, \lambda) / \mathfrak{M}_0(\lambda)$. Both tensors are smoothly defined on the punctured manifold $\mathcal{M}^*(x_0, \dots, x_k)$.

The 0th-order moment $\mathfrak{M}_0(\lambda) = \sum_i \lambda_i = \mathbf{1}^T \lambda$ is the mass. The p th-order moment is homogeneous of degree 1 in λ while the normalized p th-order moment is naturally invariant by a change of scale of the weights. For a fixed weight λ , the first-order moment $\mathfrak{M}_1(x, \lambda) = \sum_i \lambda_i \vec{x}_i$ is a smooth vector field on the manifold $\mathcal{M}^*(x_0, \dots, x_k)$ whose zeros will be the subject of our interest. The second and higher-order moments are smooth $(p, 0)$ tensor fields that will be used in contraction with the Riemannian curvature tensor.

Affinely independent points on a manifold. In a Euclidean space, $k + 1$ points are affinely independent if their affine combination generates a k dimensional subspace, or equivalently if none of the point belong to the affine span of the k others. They define in that case a k -simplex. Extending these different definitions to manifolds lead to different notions. We chose a definition which rules out the singularities of constant curvature spaces and which guaranties the existence of barycentric subspaces around reference point. In the sequel, we assume by default that the $k + 1$ reference points of pointed manifolds are affinely independent (thus $k \leq n$). Except for a few examples, the study of singular configurations is left for a future work.

DEFINITION 3 (Affinely independent points). A set of $k + 1$ points $\{x_0, \dots, x_k\}$ is affinely independent if no point is in the cut-locus of another and if all the sets of k vectors $\{\log_{x_i}(x_j)\}_{0 \leq j \neq i \leq k} \in T_{x_i} \mathcal{M}^k$ are linearly independent.

2.2. Example on the sphere \mathcal{S}_n . We consider the unit sphere in dimension $n \geq 1$ embedded in \mathbb{R}^{n+1} . The tangent space at x is the space of vectors orthogonal to x : $T_x \mathcal{S}_n = \{v \in \mathbb{R}^{n+1}, v^T x = 0\}$ and the Riemannian metric is inherited from the Euclidean metric of the embedding space. With these conventions, the Riemannian distance is the arc-length $d(x, y) = \arccos(x^T y) = \theta \in [0, \pi]$. Using the smooth function $f(\theta) = \theta / \sin \theta$ from $] -\pi; \pi[$ to \mathbb{R} , which is always greater than one, the spherical exp and log maps are

$$(3) \quad \exp_x(v) = \cos(\|v\|)x + \sin(\|v\|)v / \|v\|,$$

$$(4) \quad \log_x(y) = f(\theta)(y - \cos \theta x) \quad \text{with } \theta = \arccos(x^T y).$$

Hessian. The orthogonal projection $v = (\text{Id} - xx^T)w$ of a vector $w \in \mathbb{R}^{n+1}$ onto the tangent space $T_x \mathcal{S}_n$ provides a chart around a point $x \in \mathcal{S}_n$ where we can compute the gradient and Hessian of the squared Riemannian distance [detailed in Pennec (2017a)]. Let $u = (\text{Id} - xx^T)y / \sin \theta = \log_x(y) / \theta$ be the unit tangent vector pointing from x to y , we obtain

$$(5) \quad H_x(y) = \nabla^2 d_y^2(x) = 2uu^T + 2f(\theta) \cos \theta (\text{Id} - xx^T - uu^T).$$

By construction, x is an eigenvector with eigenvalue 0. Then the vector u [or equivalently $\log_x(y) = \theta u$] is an eigenvector with eigenvalue 1. To finish, every vector which is orthogonal to these two vectors (i.e., orthogonal to the plane spanned by $0, x$ and y) has eigenvalue $f(\theta) \cos \theta = \theta \cot \theta$. This last eigenvalue is positive for $\theta \in [0, \pi/2[$, vanishes for $\theta = \pi/2$ and becomes negative for $\theta \in]\pi/2, \pi[$. We retrieve here the results of Buss and Fillmore (2001), Lemma 2, expressed in a more general coordinate system.

Moments of a $k + 1$ -pointed sphere. We denote a set of $k + 1$ point on the sphere and the matrix of their coordinates by $X = [x_0, \dots, x_k]$. The cut locus of x_i is its antipodal point $-x_i$ so that the $(k + 1)$ -punctured manifold is $\mathcal{M}^*(x_0, \dots, x_k) = \mathcal{S}_n \setminus -X$. Using the invertible diagonal matrix $F(X, x) = \text{Diag}(f(\arccos(x_i^T x)))$, the first weighted moment is

$$(6) \quad \mathfrak{M}_1(x, \lambda) = \sum_i \lambda_i x \vec{x}_i = (\text{Id} - xx^T)XF(X, x)\lambda.$$

Affine independence of the reference points. Because no point is antipodal nor identical to another, the plane generated by $0, x_i$ and x_j in the embedding space is also generated by $0, x_i$ and the tangent vector $\log_{x_i}(x_j)$. This can be seen using a stereographic projection of pole $-x_i$ from \mathcal{S}_n to $T_{x_i}\mathcal{S}_n$. Thus, $0, x_i$ and the k independent vectors $\log_{x_i}(x_j)$ ($j \neq i$) generate the same linear subspace of dimension $k + 1$ in the embedding space than the points $\{0, x_0, \dots, x_k\}$. We conclude that $k + 1$ points on the sphere are affinely independent if and only if the matrix $X = [x_0, \dots, x_k]$ has rank $k + 1$.

2.3. *Example on the hyperbolic space \mathbb{H}^n .* We now consider the hyperboloid of equation $-x_0^2 + x_1^2, \dots, x_n^2 = -1$ ($x_0 > 0$) embedded in \mathbb{R}^{n+1} ($n \geq 2$). Using the notation $x = (x_0, \hat{x})$ and the indefinite nondegenerate symmetric bilinear form $\langle x | y \rangle_* = x^T J y = \hat{x}^T \hat{y} - x_0 y_0$ with $J = \text{diag}(-1, \text{Id}_n)$, the hyperbolic space \mathbb{H}^n can be seen as the pseudo-sphere $\|x\|_*^2 = \|\hat{x}\|^2 - x_0^2 = -1$ of radius -1 in the Minkowski space $\mathbb{R}^{1,n}$. A point can be parametrized by $x = (\sqrt{1 + \|\hat{x}\|^2}, \hat{x})$ for $\hat{x} \in \mathbb{R}^n$ (Weierstrass coordinates). The restriction of the Minkowski pseudo-metric of the embedding space $\mathbb{R}^{1,n}$ to the tangent space of $T_x \mathbb{H}^n$ is positive definite. It defines the natural Riemannian metric on the hyperbolic space. With these conventions, geodesics are the trace of 2-planes passing through the origin and the Riemannian distance is the arc-length $d(x, y) = \text{arccosh}(-\langle x | y \rangle_*)$. Using the smooth positive function $f_*(\theta) = \theta/\sinh(\theta)$ from \mathbb{R} to $]0, 1[$, the hyperbolic exp and log maps are

$$(7) \quad \exp_x(v) = \cosh(\|v\|_*)x + \sinh(\|v\|_*)v/\|v\|_*$$

$$(8) \quad \log_x(y) = f_*(\theta)(y - \cosh(\theta)x) \quad \text{with } \theta = \text{arccosh}(-\langle x | y \rangle_*).$$

Hessian. The orthogonal projection $v = w + \langle w | x \rangle_* x = (\text{Id} + xx^T J)w$ of a vector $w \in \mathbb{R}^{1,n}$ onto the tangent space at $T_x \mathbb{H}^n$ provides a chart around the point $x \in \mathbb{H}^n$ where we can compute the gradient and Hessian of the hyperbolic squared distance [detailed in Pennec (2017a)]. Let $u = \log_x(y)/\theta$ be the unit tangent vector pointing from x to y , the Hessian is

$$(9) \quad H_x(y) = \nabla^2 d_y^2(x) = 2J(uu^T + \theta \coth \theta (J + xx^T - uu^T))J$$

By construction, x is an eigenvector with eigenvalue 0. The vector u [or equivalently $\log_x(y) = \theta u$] is an eigenvector with eigenvalue 1. Every vector orthogonal to these two vectors (i.e., to the plane spanned by 0, x and y) has eigenvalue $\theta \coth \theta \geq 1$ (with equality only for $\theta = 0$). Thus, the Hessian of the squared distance is always positive definite. As a consequence, the squared distance is a convex function and has a unique minimum. This was of course expected for a negatively curved space [Bishop and O’Neill (1969)].

Moments of a $k + 1$ -pointed hyperboloid. We now pick $k + 1$ points on the hyperboloid whose matrix of coordinates is denoted by $X = [x_0, \dots, x_k]$. Since there is no cut-locus, the $(k + 1)$ -punctured manifold is the manifold itself: $\mathcal{M}^*(x_0, \dots, x_k) = \mathcal{M} = \mathbb{H}^n$. Using the invertible diagonal matrix $F_*(X, x) = \text{Diag}(f_*(\text{arccosh}(-\langle x_i | x \rangle_*)))$, the first weighted moment is

$$(10) \quad \mathfrak{M}_1(x, \lambda) = \sum_i \lambda_i \log_x(x_i) = (\text{Id} + xx^T J)X F_*(X, x)\lambda.$$

Affine independence. As for the sphere, the origin, the point x_i and the k independent vectors $\log_{x_i}(x_j) \in T_{x_i} \mathbb{H}^n$ ($j \neq i$) generate the same $k + 1$ dimensional linear subspace of the embedding Minkowski space $\mathbb{R}^{1,n}$ than the points $\{x_0, \dots, x_k\}$. Thus, $k + 1$ points on the hyperboloid are affinely independent if and only if the matrix X has rank $k + 1$.

3. Exponential Barycentric Subspaces (EBS) and affine spans.

3.1. *Affine subspaces in a Euclidean space.* In Euclidean PCA, a zero-dimensional space is a point, a one-dimensional space is a line and an affine subspace of dimension k is generated by a point and $k \leq n$ linearly independent vectors. We can also generate such a subspace by taking the affine hull of $k + 1$ affinely independent points: $\text{Aff}(x_0, \dots, x_k) = \{x = \sum_i \lambda_i x_i, \text{ with } \sum_{i=0}^k \lambda_i = 1\}$. These two definitions are equivalent in a Euclidean space, but turn out to have different generalizations in manifolds.

When there exists a vector of coefficients $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_k) \in \mathbb{R}^{k+1}$ (which do not sum to zero) such that $\sum_{i=0}^k \lambda_i (x_i - x) = 0$, then λ is called the barycentric coordinates of the point x with respect to the k -simplex $\{x_0, \dots, x_k\}$. When points are dependent, some extra care has to be taken to show that the affine span is still well defined but with a lower dimensionality. Barycentric coordinates are homogeneous of degree one.

DEFINITION 4 [Projective space of barycentric coordinates (weights)]. Barycentric coordinates of $k + 1$ points live in the real projective space $\mathbb{R}P^n = (\mathbb{R}^{k+1} \setminus \{0\})/\mathbb{R}^*$ from which we remove the codimension 1 subspace $\mathbb{1}^\perp$ orthogonal to the point $\mathbb{1} = (1 : 1 : \dots : 1)$:

$$\mathcal{P}_k^* = \{\lambda = (\lambda_0 : \lambda_1 : \dots : \lambda_k) \in \mathbb{R}P^n \text{ s.t. } \mathbb{1}^\top \lambda \neq 0\}.$$

Projective points are represented by lines through 0 in Figure 1. Standard representations are given by the intersection of the lines with the “upper” unit sphere S_k of \mathbb{R}^{k+1} with north pole $\mathbb{1}/\sqrt{k + 1}$ or by the affine k -plane of \mathbb{R}^{k+1} passing through the point $\mathbb{1}/(k + 1)$ and orthogonal to this vector. This last representation give the normalized weight $\underline{\lambda}_i = \lambda_i/(\sum_{j=0}^k \lambda_j)$: the vertices of the simplex have homogeneous coordinates $(1 : 0 : \dots : 0) \dots (0 : 0 : \dots : 1)$. To prevent weights to sum up to zero, we have to remove the codimension 1 subspace $\mathbb{1}^\perp$ orthogonal to the projective point $\mathbb{1} = (1 : 1 : \dots : 1)$ (blue line in Figure 1). This excluded subspace corresponds to the equator of the pole $\mathbb{1}/\sqrt{k + 1}$ for the sphere representation (points C and $-C$ identified in Figure 1), and to the projective completion (points at infinity) of the affine k -plane of normalized weights.

3.2. EBS and affine span in Riemannian manifolds.

DEFINITION 5 [Barycentric coordinates in a $(k + 1)$ -pointed manifold]. A point $x \in \mathcal{M}^*(x_0, \dots, x_k)$ has barycentric coordinates $\lambda \in \mathcal{P}_k^*$ with respect to

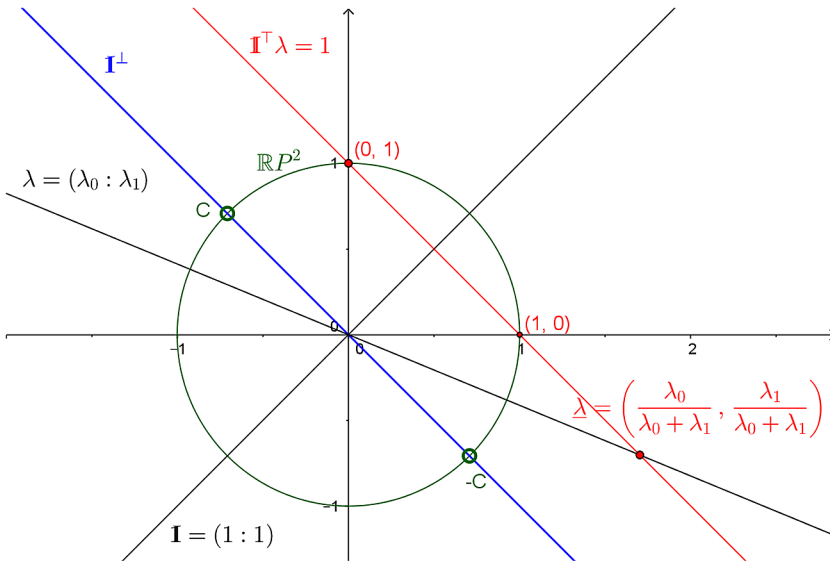


FIG. 1. Projective weights for $k = 1$.

$k + 1$ reference affinely independent points if

$$(11) \quad \mathfrak{M}_1(x, \lambda) = \sum_{i=0}^k \lambda_i \overrightarrow{xx_i} = 0.$$

Since the Riemannian log function $\overrightarrow{xx_i} = \log_x(x_i)$ is multiply defined on the cut locus of x_i , this definition cannot be extended to the union of all cut loci $C(x_0, \dots, x_k)$, which is why we restrict the definition to $\mathcal{M}^*(x_0, \dots, x_k)$.

DEFINITION 6 [Exponential Barycentric Subspace (EBS)]. The EBS of the affinely independent points $(x_0, \dots, x_k) \in \mathcal{M}^{k+1}$ is the locus of weighted exponential barycenters of the reference points in $\mathcal{M}^*(x_0, \dots, x_k)$:

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

The reference points could be seen as landmarks in the manifold. This definition is fully symmetric with respect to all of them, while one point is privileged in geodesic subspaces. We could draw a link with archetypal analysis [Cutler and Breiman (1994)] which searches for extreme data values such that all of the data can be well represented as convex mixtures of the archetypes. However, extremality is not mandatory in our framework.

PROPOSITION 1 (Dual subspace of barycentric coordinates). *The subspace of barycentric coordinates $\Lambda(x) = \{\lambda \in \mathcal{P}_k^* \mid \mathfrak{M}_1(x, \lambda) = 0\}$ at point $x \in \mathcal{M}^*(x_0, \dots, x_k)$ is either void, a point, or a linear subspace of \mathcal{P}_k^* .*

We see that a point belongs to $\text{EBS}(x_0, \dots, x_k)$ if and only if $\Lambda(x) \neq \emptyset$. Moreover, any linear combination of weights that satisfy the equation is also a valid weight so that $\Lambda(x)$ can only be a unique point (dimension 0) or a linear subspace of \mathcal{P}_k^* . The dimension of the dual space $\Lambda(x)$ is actually controlling the local dimension of the barycentric space, as we will see below.

The discontinuity of the Riemannian log on the cut locus of the reference points may hide the continuity or discontinuities of the exponential barycentric subspace. In order to ensure the completeness and potentially reconnect different components, we consider the closure of this set.

DEFINITION 7 (Affine span of $k + 1$ affinely independent points). The affine span is the closure of the EBS in \mathcal{M} : $\text{Aff}(x_0, \dots, x_k) = \overline{\text{EBS}}(x_0, \dots, x_k)$. Because we assumed that \mathcal{M} is geodesically complete, this is equivalent to the metric completion of the EBS.

3.3. *Characterizations of the EBS.* Let $Z(x) = [\vec{x}x_0, \dots, \vec{x}x_k]$ be the smooth field of $n \times (k + 1)$ matrices of vectors pointing from any point $x \in \mathcal{M}^*(x_0, \dots, x_k)$ to the reference points. We can rewrite the constraint $\sum_i \lambda_i \vec{x}x_i = 0$ in matrix form: $\mathfrak{M}_1(x, \lambda) = Z(x)\lambda = 0$, where λ is the $k + 1$ vector of homogeneous coordinates λ_i .

THEOREM 1 (Characterization of the exponential barycentric subspace). *Let $Z(x) = U(x) S(x) V(x)^T$ be a singular decomposition of the $n \times (k + 1)$ matrix field $Z(x) = [\vec{x}x_0, \dots, \vec{x}x_k]$ on $\mathcal{M}^*(x_0, \dots, x_k)$ with singular values $\{s_i(x)\}_{0 \leq i \leq k}$ sorted in decreasing order. $\text{EBS}(x_0, \dots, x_k)$ is the zero level-set of the smallest singular value $s_{k+1}(x)$ and the dual subspace of valid barycentric weights is spanned by the right singular vectors corresponding to the l vanishing singular values: $\Lambda(x) = \text{Span}(v_{k-l}, \dots, v_k)$ (it is void if $l = 0$).*

PROOF. Since U and V are orthogonal matrices, $Z(x)\lambda = 0$ if and only if at least one singular value (necessarily the smallest one s_k) is null, and λ has to live in the corresponding right-singular space: $\Lambda(x) = \text{Ker}(Z(x))$. If we have only one zero singular value ($s_{k+1} = 0$ and $s_k > 0$), then λ is proportional to v_{k+1} . If l singular values vanish, then we have a higher-dimensional linear subspace of solutions for λ . \square

THEOREM 2. *Let $G(x)$ be the matrix expression of the Riemannian metric in a local coordinate system and $\Omega(x) = Z(x)^T G(x) Z(x)$ be the smooth $(k + 1) \times (k + 1)$ matrix field on $\mathcal{M}^*(x_0, \dots, x_k)$ with components $\Omega_{ij}(x) = \langle \vec{x}x_i | \vec{x}x_j \rangle_x$ and $\Sigma(x) = \mathfrak{M}_2(x, \mathbb{1}) = \sum_{i=0}^k \vec{x}x_i \vec{x}x_i^T = Z(x) Z(x)^T$ be the (scaled) $n \times n$ covariance matrix field of the reference points. $\text{EBS}(x_0, \dots, x_k)$ is the zero level-set of: $\det(\Omega(x))$, the minimal eigenvalue σ_{k+1}^2 of $\Omega(x)$, the $k + 1$ eigenvalue (in decreasing order) of the covariance $\Sigma(x)$.*

PROOF. The constraint $\mathfrak{M}_1(x, \lambda) = 0$ is satisfied if and only if:

$$\|\mathfrak{M}_1(x, \lambda)\|_x^2 = \left\| \sum_i \lambda_i \vec{x}x_i \right\|_x^2 = \lambda^T \Omega(x) \lambda = 0.$$

As the function is homogeneous in λ , we can restrict to unit vectors. Adding this constrains with a Lagrange multiplier to the cost function, we end-up with the Lagrangian $\mathcal{L}(x, \lambda, \alpha) = \lambda^T \Omega(x) \lambda + \alpha(\lambda^T \lambda - 1)$. The minimum with respect to λ is obtained for the eigenvector $\mu_{k+1}(x)$ associated to the smallest eigenvalue $\sigma_{k+1}(x)$ of $\Omega(x)$ (assuming that eigenvalues are sorted in decreasing order) and we have $\|\mathfrak{M}_1(x, \mu_{k+1}(x))\|_2^2 = \sigma_{k+1}(x)$, which is null if and only if the minimal eigenvalue is zero. Thus, the barycentric subspace of $k + 1$ points is the locus of rank deficient matrices $\Omega(x)$:

$$\text{EBS}(x_0, \dots, x_k) = \phi^{(-1)}(0) \quad \text{where } \phi(x) = \det(\Omega(x)).$$

One may want to relate the singular values of $Z(x)$ to the eigenvalues of $\Omega(x)$. The later are the square of the singular values of $G(x)^{1/2}Z(x)$. However, the left multiplication by the square root of the metric (a nonsingular but nonorthogonal matrix) obviously changes the singular values in general except for vanishing ones: the (right) kernels of $G(x)^{1/2}Z(x)$ and $Z(x)$ are indeed the same. This shows that the EBS is an affine notion rather than a metric one, contrarily to the Fréchet/Karcher barycentric subspace.

To draw the link with the $n \times n$ covariance matrix of the reference points, let us notice first that the definition does not assumes that the coordinate system is orthonormal. Thus, the eigenvalues of the covariance matrix depend on the chosen coordinate system, unless they vanish. In fact, only the joint eigenvalues of $\Sigma(x)$ and $G(x)$ really make sense, which is why this decomposition is called the proper orthogonal decomposition (POD). Now, the singular values of $Z(x) = U(x)S(x)V(x)^T$ are also the square root of the first $k + 1$ eigenvalues of $\Sigma(x) = U(x)S^2(x)U(x)^T$, the remaining $n - k - 1$ eigenvalues being null. Similarly, the singular values of $G(x)^{1/2}Z(x)$ are the square root of the first $k + 1$ joint eigenvalues of $\Sigma(x)$ and $G(x)$. Thus, our barycentric subspace may also be characterized as the zero level-set of the $k + 1$ eigenvalue (sorted in decreasing order) of Σ , and this characterization is once again independent of the basis chosen. \square

3.4. *Spherical EBS and affine span.* From equation (6), we identify the matrix: $Z(x) = (\text{Id} - xx^T)XF(X, x)$. Finding points x and weights λ such that $Z(x)\lambda = 0$ is a classical matrix equation, except for the scaling matrix $F(X, x)$ acting on homogeneous projective weights, which is nonstationary and nonlinear in both X and x . However, since $F(X, x) = \text{Diag}(\theta_i / \sin \theta_i)$ is an invertible diagonal matrix, we can introduce *renormalized weights* $\tilde{\lambda} = F(X, x)\lambda$, which leaves us with the equation $(\text{Id} - xx^T)X\tilde{\lambda} = 0$. The solutions under the constraint $\|x\| = 1$ are given by $(x^T X\tilde{\lambda})x = X\tilde{\lambda}$ or more explicitly $x = \pm X\tilde{\lambda} / \|X\tilde{\lambda}\|$ whenever $X\tilde{\lambda} \neq 0$. This condition is ensured if $\text{Ker}(X) = \{0\}$. Thus, when the reference points are linearly independent, the point $x \in \mathcal{M}^*(X)$ has to belong to the Euclidean span of the reference vectors. Notice that for each barycentric coordinate we have two antipodal solution points. Conversely, any unit vector $x = X\tilde{\lambda}$ of the Euclidean span of X satisfies the equation $(\text{Id} - xx^T)X\tilde{\lambda} = (1 - \|x\|^2)X\tilde{\lambda} = 0$, and is thus a point of the EBS provided that it is not at the cut-locus of one of the reference points. This shows that

$$(12) \quad \text{EBS}(X) = \text{Span}\{x_0, \dots, x_k\} \cap \mathcal{S}_n \setminus X.$$

Using the renormalization principle, we can orthogonalize the reference points: let $X = USV^T$ be a singular value decomposition of the matrix of reference vectors. All the singular values s_i are positive since the reference vectors x_i are assumed to be linearly independent. Thus, $\mu = SV^T\tilde{\lambda} = SV^TF(X, x)\lambda$ is an invertible change of coordinate, and we are left with solving $(\text{Id} - xx^T)U\mu = 0$. By definition of the singular value decomposition, the Euclidean spans of X and U are the

same, so that $EBS(U) = \text{Span}\{x_0, \dots, x_k\} \cap \mathcal{S}_n \setminus -U$. This shows that the exponential barycentric subspace generated by the original points $X = [x_0, \dots, x_k]$ and the orthogonalized points $U = [u_0, \dots, u_k]$ are the same, except at the cut locus of all these points, but with different barycentric coordinates.

To obtain the affine span, we take the closure of the EBS, which incorporates the cut locus of the reference points: $\text{Aff}(X) = \text{Span}\{x_0, \dots, x_k\} \cap \mathcal{S}_n$. Thus, for spherical data as for Euclidean data, the affine span only depend on the reference points through the point of the Grassmanian they define.

THEOREM 3 (Spherical affine span). *The affine span $\text{Aff}(X)$ of $k + 1$ linearly independent reference unit points $X = [x_0, \dots, x_k]$ on the n -dimensional sphere \mathcal{S}_n endowed with the canonical metric is the great subsphere of dimension k that contains the reference points.*

When the reference points are affinely dependent on the sphere, the matrix X has one or more (say l) vanishing singular values. Any weight $\tilde{\lambda} \in \text{Ker}(X)$ is a barycentric coordinate vector for any point x of the pointed sphere since the equation $(\text{Id} - xx^T)X\tilde{\lambda} = 0$ is verified. Thus, the EBS is $\mathcal{S}_n \setminus -X$ and the affine span is the full sphere. If we exclude the abnormal subspace of weights valid for all points, we find that x should be in the span of the nonzero left singular vectors of X , that is, in the subsphere of dimension of dimension $\text{rank}(X) - 1$ generated the Euclidean span of the reference vectors. This can also be achieved by focusing of the locus of points where $Z(x)$ has two vanishing singular values. This more reasonable result suggests adapting the EBS and affine span definitions for singular point configurations.

Two points on a 2-sphere is an interesting example that can be explicitly worked out. When the points are not antipodal, the rank of $X = [x_0, x_1]$ is 2, and the generated affine span is the one-dimensional geodesic joining the two points. When the reference points are antipodal, say north and south poles, X becomes rank one and one easily sees that all points of the 2-sphere are on one geodesic joining the poles with opposite log directions to the poles. This solution of the EBS definition correspond to the renormalized weight $\tilde{\lambda} = (1/2 : 1/2) \in \text{Ker}(X)$ of the kernel of X . However, looking at the locus of points with two vanishing singular values of $Z(x)$ leads to restrict to the north and south poles only, which is a more natural and expected result.

3.5. Hyperbolic EBS and affine span. The hyperbolic case closely follows the spherical one. From equation (10), we get the expression of the matrix $Z(x) = (\text{Id} + xx^T J)X F_*(X, x)$. Solving for $Z(x)\lambda = 0$ can be done as previously by solving $(\text{Id} + xx^T J)X\tilde{\lambda} = 0$ with the renormalized weights $\tilde{\lambda} = F_*(X, x)\lambda$. This equation rewrites $\langle x | X\tilde{\lambda} \rangle_* x = -X\tilde{\lambda}$, so that the solution has to be of the form $X\tilde{\lambda} = 0$ or $x = \alpha X\tilde{\lambda}$. When the points are affinely independent, the first form is excluded since $\text{Ker}(X) = 0$. In order to satisfy the constraint $\|x\|_*^2 = -1$ in the second form,

we need to have $\alpha^2 = -\|X\tilde{\lambda}\|_*^{-2} > 0$ and the first coordinate $[X\tilde{\lambda}]_0$ of $X\tilde{\lambda}$ has to be positive. This defines a cone in the space of renormalized weights from which each line parametrizes a point $x = \text{sgn}([X\tilde{\lambda}]_0)X\tilde{\lambda}/\sqrt{-\|X\tilde{\lambda}\|_*^2}$ of the Hyperbolic EBS. Thus, $\text{Aff}(X)$ is the k -dimensional hyperboloid generated by the intersection of the Euclidean span of the reference vectors with the hyperboloid \mathbb{H}^n . Since it is complete, the completion does not add anything to the affine span:

$$(13) \quad \text{Aff}(X) = \text{EBS}(X) = \text{Span}\{x_0, \dots, x_k\} \cap \mathbb{H}^n.$$

As for spheres, we see that the hyperbolic affine span only depend on the reference points through the point of the Grassmanian they define.

THEOREM 4 (Hyperbolic affine span). *The affine span $\text{Aff}(X) = \text{EBS}(X)$ of $k + 1$ affinely independent reference points $X = [x_0, \dots, x_k]$ on the n -dimensional hyperboloid \mathbb{H}^n endowed with the canonical Minkowski pseudo-metric of the embedding space $\mathbb{R}^{1,n}$ is the hyperboloid of dimension k generated by the intersection of the hyperboloid with the hyperplane containing the reference points.*

When the matrix X has one or more vanishing singular values (affine dependence), all the points of the hyperboloid are solutions corresponding to weights from $\text{Ker}(X)$. Excluding these abnormal solutions and looking at the locus of points where $Z(x)$ has two vanishing singular values, we find that x should be in the span of the nonzero left singular vectors of X , that is, in the sub-hyperboloid of dimension $\text{rank}(X) - 1$ generated the Euclidean span of the reference vectors.

4. Fréchet/Karcher barycentric subspaces. The reformulation of the affine span as the weighted mean of $k + 1$ points also suggests a definition using the Fréchet or the Karcher mean, valid in general metric spaces.

DEFINITION 8 (Fréchet/Karcher barycentric subspaces of $k + 1$ points). Let $(\mathcal{M}, \text{dist})$ be a metric space of dimension n and $(x_0, \dots, x_k) \in \mathcal{M}^{k+1}$ be $k + 1 \leq n + 1$ distinct reference points. The (normalized) weighted variance at point x with weight $\lambda \in \mathcal{P}_k^*$ is: $\sigma^2(x, \lambda) = \frac{1}{2} \sum_{i=0}^k \lambda_i \text{dist}^2(x, x_i) = \frac{1}{2} \sum_{i=0}^k \lambda_i \text{dist}^2(x, x_i) / (\sum_{j=0}^k \lambda_j)$. The Fréchet barycentric subspace of these points is the locus of weighted Fréchet means of these points, that is, the set of absolute minima of the weighted variance:

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

The Karcher barycentric subspaces $\text{KBS}(x_0, \dots, x_k)$ are defined similarly with local minima instead of global ones.

In stratified metric spaces, for instance, the barycentric subspace spanned by points belonging to different strata naturally maps over several strata. This is a

significant improvement over geodesic subspaces used in PGA which can only be defined within a regular strata. In the sequel, we only deal with the KBS/FBS of affinely independent points in a Riemannian manifold.

4.1. *Link between the different barycentric subspaces.* In order to analyze the relationship between the Fréchet, Karcher and Exponential barycentric subspaces, we follow the seminal work of [Karcher \(1977\)](#). First, the locus of local minima (i.e., Karcher mean) is a superset of the global minima (Fréchet mean). On the punctured manifold $\mathcal{M}^*(x_0, \dots, x_k)$, the squared distance $d_{x_i}^2(x) = \text{dist}^2(x, x_i)$ is smooth and its gradient is $\nabla d_{x_i}^2(x) = -2 \log_x(x_i)$. Thus, one recognizes that the EBS equation $\sum_i \lambda_i \log_x(x_i) = 0$ [equation (11)] defines nothing else than the critical points of the weighted variance:

$$\text{FBS} \cap \mathcal{M}^* \subset \text{KBS} \cap \mathcal{M}^* \subset \text{Aff} \cap \mathcal{M}^* = \text{EBS}.$$

Among the critical points with a nondegenerate Hessian, local minima are characterized by a positive definite Hessian. When the Hessian is degenerate, we cannot conclude on the local minimality without going to higher-order differentials. The goal of this section is to subdivide the EBS into a cell complex according to the index of the Hessian operator of the variance:

$$(14) \quad H(x, \lambda) = \nabla^2 \sigma^2(x, \lambda) = - \sum_{i=0}^k \lambda_i D_x \log_x(x_i).$$

Plugging the value of the Taylor expansion of the differential of the log of equation (1), we obtain the Taylor expansion:

$$(15) \quad [H(x, \lambda)]_b^a = \delta_b^a - \frac{1}{3} R_{cbd}^a(x) \mathfrak{M}_2^{cd}(x, \underline{\lambda}) - \frac{1}{12} \nabla_c R_{dbe}^a(x) \mathfrak{M}_3^{cde}(x, \underline{\lambda}) + O(\varepsilon^4).$$

The key factor in this expression is the contraction of the Riemannian curvature with the weighted covariance tensor of the reference points. This contraction is an extension of the Ricci curvature tensor. Exactly as the Ricci curvature tensor encodes how the volume of an isotropic geodesic ball in the manifold deviates from the volume of the standard ball in a Euclidean space (through its metric trace, the scalar curvature), the extended Ricci curvature encodes how the volume of the geodesic ellipsoid $\vec{x}\vec{y}^T \mathfrak{M}_2(x, \underline{\lambda})^{(-1)} \vec{x}\vec{y} \leq \varepsilon$ deviates from the volume of the standard Euclidean ellipsoid.

In locally symmetric affine spaces, the covariant derivative of the curvature is identically zero, which simplifies the formula. In the limit of null curvature (e.g., for a locally Euclidean space like the torus), the Hessian matrix $H(x, \lambda)$ converges to the unit matrix and never vanishes. In general, Riemannian manifolds, equation (15) only gives a qualitative behavior but does not provide guaranties as it is a

series involving higher-order moments of the reference points. In order to obtain hard bounds on the spectrum of $H(x, \lambda)$, one has to investigate bounds on Jacobi fields using Riemannian comparison theorems, as for the proof of uniqueness of the Karcher and Fréchet means [see Karcher (1977), Kendall (1990), Le (2004), Afsari (2011), Yang (2011)].

DEFINITION 9 (Degenerate, nondegenerate and positive points). An exponential barycenter $x \in \text{EBS}(x_0, \dots, x_k)$ is degenerate (resp., nondegenerate or positive) if the Hessian matrix $H(x, \lambda)$ is singular (resp., definite or positive definite) for all λ in the dual space of barycentric coordinates $\Lambda(x)$. The set of degenerate exponential barycenters is denoted by $\text{EBS}^0(x_0, \dots, x_k)$ [resp., nondegenerate by $\text{EBS}^*(x_0, \dots, x_k)$ and positive by $\text{EBS}^+(x_0, \dots, x_k)$].

The definition of nondegenerate and positive points could be generalized to non-critical points (outside the affine span) by considering for instance the right singular space of the smallest singular value of $Z(x)$. However, this would depend on the metric on the space of weights and a renormalization of the weights (such as for spheres) can change the smallest nonzero singular value. Positive points are obviously nondegenerate. In Euclidean spaces, all the points of an affine span are positive and nondegenerate. In positively curved manifolds, we may have degenerate points and nonpositive points, as we will see with the sphere example. For negatively curved spaces, the intuition that points of the EBS should all be positive like in Euclidean spaces is also wrong, as we will see with the example of hyperbolic spaces.

THEOREM 5 (Karcher barycentric subspace and positive span). $\text{EBS}^+(x_0, \dots, x_k)$ is the set of nondegenerate points of the Karcher barycentric subspace $\text{KBS}(x_0, \dots, x_k)$ on $\mathcal{M}^*(x_0, \dots, x_k)$. In other words, 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.

4.2. Spherical KBS. In order to find the positive points of the EBS on the sphere, we compute the Hessian of the normalized variance. Using equation (5) and $u_i = \log_x(x_i)/\theta_i$, we obtain the Hessian of $\sigma^2(x, \lambda) = \frac{1}{2} \sum_{i=0}^k \lambda_i \text{dist}^2(x, x_i)$:

$$H(x, \lambda) = \left(\sum_i \lambda_i \theta_i \cot \theta_i \right) (\text{Id} - x x^T) + \sum_i \lambda_i (1 - \theta_i \cot \theta_i) u_i u_i^T.$$

As expected, x is an eigenvector with eigenvalue 0 due to the projection on the tangent space at x . Any vector w of the tangent space at x (thus orthogonal to x) which is orthogonal to the affine span (and thus to the vectors u_i) is an eigenvector with eigenvalue $\sum_i \lambda_i \theta_i \cot \theta_i$. Since the Euclidean affine span $\text{Aff}_{\mathbb{R}^{n+1}}(X)$ has $\text{rank}(X) \leq k + 1$ dimensions, this eigenvalue has multiplicity

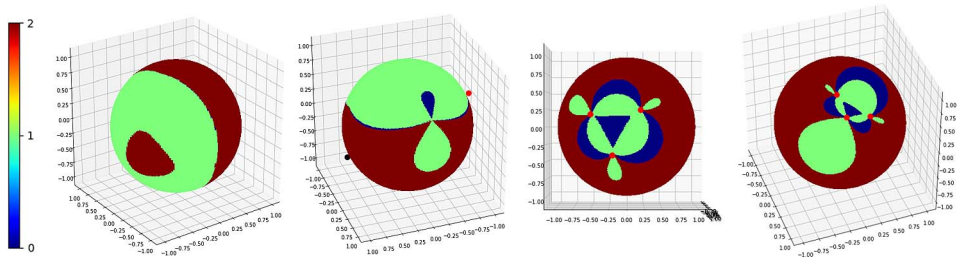


FIG. 2. Signature of the weighted Hessian matrix for different configurations of 3 reference points (in black, antipodal point in red) on the 2-sphere: the locus of local minima (KBS) in brown does not cover the whole sphere and can even be disconnected (first example).

$n + 1 - \text{rank}(X) \geq n - k$ when $x \in \text{Aff}(X)$. The last $\text{Rank}(X) - 1$ eigenvalues have associated eigenvectors within $\text{Aff}_{\mathbb{R}^{n+1}}(X)$.

Buss and Fillmore (2001) have shown that this Hessian matrix is positive definite for *positive weights* when the points are within one hemisphere with at least one nonzero weight point which is not on the equator. In contrast, we are interested here in the positivity and definiteness of the Hessian $H(x, \lambda)$ for the positive and negative weights which live in dual space of barycentric coordinates $\Lambda(x)$. This is actually a nontrivial algebraic geometry problem. Simulation tests with random reference points X show that the eigenvalues of $H(x, \underline{\lambda}(x))$ can be positive or negative at different points of the EBS. The number of positive eigenvalues (the index) of the Hessian is illustrated on Figure 2 for a few configurations of 3 affinely independent reference points on the 2-sphere. This illustrates the subdivision of the EBS on spheres in a cell complex based on the index of the critical point: the positive points of the KBS do not in general cover the full subsphere containing the reference points. It may even be disconnected, contrarily to the affine span which consistently covers the whole subsphere. For subspace definition purposes, this suggests that the affine span might thus be the most interesting definition. For affinely dependent points, the KBS/FBS behave similar to the EBS. For instance, the weighted variance of $X = [e_1, -e_1]$ on a 2-sphere is a function of the latitude only. The points of a parallel at any specific latitude are global minima of the weighted variance for a choice of $\lambda = (\alpha : 1 - \alpha)$, $\alpha \in [0, 1]$. Thus, all points of the sphere belong to the KBS, which is also the FBS and the affine span. However, the Hessian matrix has one positive eigenvalue along meridians and one zero eigenvalue along the parallels. This is a very nongeneric case.

4.3. *Hyperbolic KBS/FBS.* Let $x = X\tilde{\lambda}$ be a point of the hyperbolic affine span of $X = [x_0, \dots, x_k]$. The renormalized weights $\tilde{\lambda}$ are related to the original weights through $\lambda = F_*(X, x)^{-1}\tilde{\lambda}$ and satisfy $\|X\tilde{\lambda}\|_*^2 = -1$ and $\text{sgn}([X\tilde{\lambda}]_0) > 0$. The point x is a critical point of the (normalized) weighted variance. In order to know if this is a local minimum (i.e., a point of the KBS), we compute the Hessian of this weighted variance. Denoting $u_i = \log_x(x_i)/\theta_i$ with $\cosh \theta_i = -\langle x | x_i \rangle_*$,

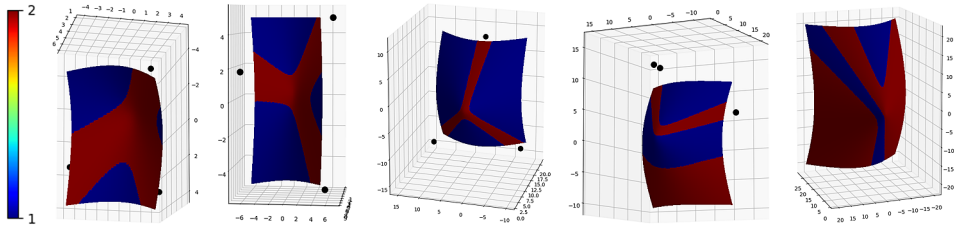


FIG. 3. Signature of the weighted Hessian matrix for different configurations of 3 reference points on the 2-hyperboloid: the locus of local minima (KBS) in brown does not cover the whole hyperboloid and can be disconnected (last two examples).

and using the Hessian of the square distance derived in equation (10), we obtain the following formula:

$$H(x, \lambda) = \sum_i \underline{\lambda}_i \theta_i \coth \theta_i (J + Jxx^T J) + \sum_i \underline{\lambda}_i (1 - \theta_i \coth \theta_i) J u_i u_i^T J.$$

As expected, x is an eigenvector with eigenvalue 0 due to the projection on the tangent space at x . Any vector w of the tangent space at x , which is orthogonal to the affine span (and thus to the vectors u_i), is an eigenvector with eigenvalue $\sum_i \underline{\lambda}_i \theta_i \coth \theta_i = 1/(\mathbb{1}^T \tilde{\lambda})$ with multiplicity $n + 1 - \text{rank}(X)$. The last $\text{Rank}(X) - 1$ eigenvalues have associated eigenvectors within $\text{Aff}_{\mathbb{R}^{n+1}}(X)$. Simulation tests with random reference points X show these eigenvalues can be positive or negative at different points of $\text{Aff}(X)$. The index of the Hessian is illustrated on Figure 3 for a few configuration of 3 affinely independent reference points on the 2-hyperbolic space. Contrarily to the sphere, we observe only one or two positive eigenvalues corresponding respectively to saddle points and local minima. This subdivision of the hyperbolic affine span in a cell complex shows that the hyperbolic KBS is in general a strict subset of the hyperbolic affine span. We conjecture that there is an exception for reference points at infinity, for which the barycentric subspaces could be generalized using Busemann functions [Busemann (1955)]: it is likely that the FBS, KBS and the affine span are all equal in this case and cover the whole lower dimensional hyperbola.

5. Properties of the barycentric subspaces. The EBS exists at each reference point x_i with weight 1 for this point and zero for the others. Moreover, when the points are affinely independent, the matrix $Z(x_i)$ has exactly one zero singular value since column i is $\log_{x_i}(x_i) = 0$ and all the other column vectors are affinely independent. Finally, the weighted Hessian matrix boils down to $H(x_i, \lambda) = -D_x \log_x(x_i)|_{x=x_i} = \text{Id}$ [see, e.g., equation (1)]. Thus, the reference points are actually local minima of the weighted variance and the KBS exists by continuity in their neighborhood.

5.1. *Barycentric simplex in a regular geodesic ball.* We call the subset of the FBS that has nonnegative weights a barycentric simplex. It contains all the reference points, the geodesics segments between the reference points, and of course the Fréchet mean of the reference points. This is the generalization of a geodesic segment for 2 points, a triangle for 3 points, etc. The $(k - l)$ -faces of a k -simplex are the simplices defined by the barycentric subspace of $k - l + 1$ points among the $k + 1$. They are obtained by imposing the l remaining barycentric coordinates to be zero. In parallel to this paper, [Weyenberg \(2015\)](#) has investigated barycentric simplexes as extensions of principal subspaces in the negatively curved metric spaces of trees under the name Locus of Fréchet mean (LFM), with very interesting results.

THEOREM 6 (Barycentric simplex in a regular geodesic ball). *Let κ be an upper bound of sectional curvatures of \mathcal{M} and $\text{inj}(\mathcal{M})$ be the radius of injection (which can be infinite) of the Riemannian manifold. Let $X = \{x_0, \dots, x_k\} \in \mathcal{M}^{(k+1)}$ be a set of $k + 1 \leq n$ affinely independent points included in a regular geodesic ball $B(x, \rho)$ with $\rho < \frac{1}{2} \min\{\text{inj}(\mathcal{M}), \frac{1}{2}\pi/\sqrt{\kappa}\}$ ($\pi/\sqrt{\kappa}$ being infinite if $\kappa < 0$). The barycentric simplex is the graph of a k -dimensional differentiable function from the nonnegative quadrant of homogeneous coordinates $(\mathcal{P}_k^*)^+$ to $B(x, \rho)$ and is thus at most k -dimensional. The $(k - l)$ -faces of the simplex are the simplices defined by the barycentric subspace of $k - l + 1$ points among the $k + 1$ and include the reference points themselves as vertices and the geodesics joining them as edges.*

PROOF. The proof closely follows the one of [Karcher \(1977\)](#) for the uniqueness of the Riemannian barycenter. The main argument is that $\mu_{(X,\lambda)}(x) = \sum \underline{\lambda}_i \delta_{x_i}(x)$ is a probability distribution whose support is included in the strongly convex geodesic ball $B(x, \rho)$. The variance $\sigma^2(x, \lambda) = \frac{1}{2} \sum_i \underline{\lambda}_i d^2(x, x_i)$ is strictly convex on that ball and has a unique minimum $x_\lambda \in B(x, \rho)$, necessarily the weighted Fréchet mean. This proof of the uniqueness of the weighted Fréchet mean with nonnegative weights was actually already present in [Buser and Karcher \(1981\)](#). We supplement the proof here by noting that since the Hessian $H(x_\lambda, \lambda) = \sum_i \underline{\lambda}_i H_i(x_\lambda)$ is the convex combination of positive matrices, it is positive definite for all $\lambda \in (\mathcal{P}_k^*)^+$ in the positive quadrant. Thus the function x_λ is differentiable thanks to the implicit function theorem: $D_\lambda x_\lambda = H(x_\lambda, \lambda)^{(-1)} Z(x_\lambda)$. The rank of this derivative is at most k since $Z(x_\lambda) = 0$, which proves that the graph of the function x_λ describes at most a k dimensional subset in \mathcal{M} . \square

5.2. *Barycentric simplexes and convex hulls.* In a vector space, a point lies in the convex hull of a simplex if and only if its barycentric coordinates are all nonnegative (thus between 0 and 1 with the unit sum constraint). Consequently, barycentric coordinates are often thought to be related to convex hulls. However,

in a general Riemannian manifold, the situation is quite different. When there are closed geodesics, the convex hull can reveal several disconnected components, unless one restrict to convex subsets of the manifolds as shown by Groisser (2004). In metric spaces with negative curvature (CAT spaces), Weyenberg (2015) displays explicit examples of convex hulls of 3 points which are 3-dimensional rather than 2-dimensional as expected. In fact, the relationship between barycentric simplexes and convex hulls cannot hold in general Riemannian manifolds if the barycentric simplex is not totally geodesic at each point, which happens for constant curvature spaces but not for general Riemannian manifolds.

5.3. Local dimension of the barycentric subspaces. Let x be a point of the EBS with affinely independent reference points. The EBS equation $Z(x)\lambda = 0$ for $\lambda \in \Lambda(x)$ is smooth in x and λ so that we can take a Taylor expansion: at the first order, a variation of barycentric coordinates $\delta\lambda$ induces a variation of position δx which are linked through $H(x, \lambda)\delta x - Z(x)\delta\lambda = 0$. Thus, at regular points:

$$\delta x = H(x, \lambda)^{(-1)}Z(x)\delta\lambda.$$

Let $Z(x) = U(x)S(x)V(x)^T$ be a singular value decomposition with singular values sorted in decreasing order. Since x belongs to the EBS, there is at least one (say $m \geq 1$) singular value that vanish and the dual space of barycentric coordinates is $\Lambda(x) = \text{Span}(v_{k-m}, \dots, v_k)$. For a variation of weights $\delta\lambda$ in this subspace, there is no change of coordinates, while any variation of weights in $\text{Span}(v_0, \dots, v_{k-m-1})$ induces a nonzero position variation. Thus, the tangent space of the EBS restricts to the $(k - m)$ -dimensional linear space generated by $\{\delta x'_i = H(x, \lambda)^{(-1)}u_i\}_{0 \leq i \leq k-m}$. Here, we see that the Hessian matrix $H(x, \lambda)$ encodes the distortion of the orthonormal frame field $u_1(x), \dots, u_k(x)$ to match the tangent space. Since the lower dimensional subspaces are included one the larger ones, we have a stratification of our k -dimensional submanifold into $k - 1, k - 2, \dots, 0$ -dimensional subsets.

THEOREM 7 (Dimension of the exponential barycentric subspace at nondegenerate points). *The nondegenerate exponential barycentric subspace $\text{EBS}^*(x_0, \dots, x_k)$ of $k + 1$ affinely independent points is a stratified space of dimension k on $\mathcal{M}^*(x_0, \dots, x_k)$. On the m -dimensional strata, $Z(x)$ has exactly $k - m + 1$ vanishing singular values.*

At degenerate points, $H(x, \lambda)$ is not invertible and vectors living in its kernel are also authorized, which potentially raises the dimensionality of the tangent space, even if they do not change the barycentric coordinates. These pathologies do not appear in practice for the constant curvature spaces as we have seen with spherical and hyperbolic spaces, and we conjecture that this is also not the case for symmetric spaces.

5.4. *Stability of the affine span with respect to the metric power.* The Fréchet (resp., Karcher) mean can be further generalized by taking a power p of the metric to define the p -variance $\sigma^p(x) = \frac{1}{p} \sum_{i=0}^k \text{dist}^p(x, x_i)$. The global (resp., local) minima of this p -variance defines the median for $p = 1$. This suggest to further generalize barycentric subspaces by taking the locus of the minima of the weighted p -variance $\sigma^p(x, \lambda) = \frac{1}{p} \sum_{i=0}^k \lambda_i \text{dist}^p(x, x_i)$. In fact, it turns out that all these “ p -subspaces” are necessarily included in the affine span, which shows that this notion is really central. To see that, we compute the gradient of the p -variance at nonreference point of $\mathcal{M}^*(x_0, \dots, x_k)$:

$$\nabla_x \sigma^p(x, \lambda) = - \sum_{i=0}^k \lambda_i \text{dist}^{p-2}(x, x_i) \log_x(x_i).$$

Critical points of the p -variance satisfy the equation $\sum_{i=0}^k \lambda'_i \log_x(x_i) = 0$ for the new weights $\lambda'_i = \lambda_i \text{dist}^{p-2}(x, x_i)$. Thus, they are still elements of the EBS and changing the power of the metric just amounts to a reparametrization of the barycentric weights.

5.5. *Restricted geodesic submanifolds are limit of affine spans.* We investigate in this section what is happening when all the points $\{x_i = \exp_{x_0}(\varepsilon w_i)\}_{1 \leq i \leq k}$ are converging to x_0 at first order along k independent vectors $\{w_i\}_{1 \leq i \leq k}$. Here, we fix $w_0 = 0$ to simplify the derivations, but the proof can be easily extended with a suitable change of coordinates provided that $\sum_{i=0}^k w_i = 0$. In Euclidean spaces, a point of the affine span $y = \sum_{i=0}^k \lambda_i x_i$ may be written as the point $y = x + \varepsilon \sum_{i=1}^k \lambda_i w_i$ of the “geodesic subspace” generated by the family of vectors $\{w_i\}_{1 \leq i \leq k}$. By analogy, we expect the exponential barycentric subspace $\text{EBS}(x_0, \exp_{x_0}(\varepsilon w_1), \dots, \exp_{x_0}(\varepsilon w_k))$ to converge towards the totally geodesic subspace at x generated by the k independent vectors w_1, \dots, w_k of $T_x \mathcal{M}$:

$$\text{GS}(x, w_1, \dots, w_k) = \left\{ \exp_x \left(\sum_{i=1}^k \alpha_i w_i \right) \in \mathcal{M} \text{ for } \alpha \in \mathbb{R}^k \right\}.$$

In fact, the above definition of the geodesic subspaces (which is the one implicitly used in most of the works using PGA) is too large and may not define a k -dimensional submanifold when there is a cut-locus. For instance, it is well known that geodesics of a flat torus are either periodic or everywhere dense in a flat torus submanifold depending on whether the components of the initial velocity field have rational or irrational ratios. This means that the geodesic space generated by a single vector for which all ratio of coordinates are irrational [e.g., $w = (\pi, \pi^2, \dots, \pi^k)$] is filling the full k -dimensional flat torus. Thus, all the 1-dimensional geodesic subspaces that have irrational ratio of coordinates minimize the distance to any set of data points in a flat torus of any dimension. In order to have a more meaningful definition and to guaranty the dimensionality of the

geodesic subspace, we need to restrict the definition to the points of the geodesics that are distance minimizing.

DEFINITION 10 (Restricted geodesic submanifolds). Let $x \in \mathcal{M}$ be a point of a Riemannian manifold and let $W_x = \{\sum_{i=1}^k \alpha_i w_i, \alpha \in \mathbb{R}^k\}$ be the k -dimensional linear subspace of $T_x \mathcal{M}$ generated a k -tuple $\{w_i\}_{1 \leq i \leq k} \in (T_x \mathcal{M})^k$ of independent tangent vectors at x . We consider the geodesics starting at x with tangent vectors in W_x , but up to the first cut-point of x only. This generates a submanifold of \mathcal{M} called the restricted geodesic submanifold $\text{GS}^*(W_x)$:

$$\text{GS}^*(W_x) = \text{GS}^*(x, w_1, \dots, w_k) = \{\exp_x(w), w \in W_x \cap D(x)\},$$

where $D(x) \subset T_x \mathcal{M}$ is the injectivity domain.

It may not be immediately clear that the subspace we define that way is a submanifold of \mathcal{M} : since \exp_x is a diffeomorphism from $D(x) \subset T_x \mathcal{M}$ to $\mathcal{M} \setminus \mathcal{C}(x)$ whose differential has full rank, its restriction to the open star-shape subset $W_x \cap D(x)$ of dimension k is a diffeomorphism from that subset to the restricted geodesic subspace $\text{GS}^*(W_x)$ which is thus an open submanifolds of dimension k of \mathcal{M} . This submanifold is generally not geodesically complete.

THEOREM 8 (Restricted geodesic subspaces are limit of affine spans). *The restricted geodesic submanifold $\text{GS}^*(W_{x_0}) = \{\exp_{x_0}(w), w \in W_{x_0} \cap D(x_0)\}$ is the limit of the $\text{EBS}(x_0, x_1(\varepsilon), \dots, x_k(\varepsilon))$ when the points $x_i(\varepsilon) = \exp_{x_0}(\varepsilon w_i)$ are converging to x_0 at first order in ε along the tangent vectors w_i defining the k -dimensional subspace $W_{x_0} \subset T_{x_0} \mathcal{M}$. These limit points are parametrized by barycentric coordinates at infinity in the codimension 1 subspace $\mathbb{1}^\perp$, the projective completion of \mathcal{P}_k^* in $\mathbb{R}P^k$; see Definition 4.*

The proof is deferred to the [Appendix](#) because of its technicality. We conjecture that the construction can be generalized using techniques from sub-Riemannian geometry to higher-order derivatives when the first-order derivative do not span a k -dimensional subspace. This would mean that we could also see some nongeodesic decomposition schemes as limit cases of barycentric subspaces, such as splines on manifolds [[Crouch and Leite \(1995\)](#), [Machado, Silva Leite and Krakowski \(2010\)](#), [Gay-Balmaz et al. \(2012\)](#)].

Example on spheres and hyperbolic spaces. In spheres (resp., hyperbolic spaces), the restricted geodesic subspace $\text{GS}^*(W_x)$ describes a great subsphere (resp., a great hyperbola), except for the cut-locus of the base-point x in spheres. Thus, points of $\text{GS}^*(W_x)$ are also points of the affine span generated by $k + 1$ affinely independent reference points of this subspace. When all the reference points $x_i = \exp_x(\varepsilon w_i)$ coalesce to a single point x along the tangent vectors

$W = [w_0, \dots, w_k]$ (with $W\mathbb{1} = 0$), we find that solutions of the EBS equation are of the form $y = x + W(\varepsilon\tilde{\lambda}/\mathbb{1}^T\tilde{\lambda}) + O(\varepsilon^2)$, which describes the affine hyperplane generated by x and W in the embedding Euclidean (resp., Minkowski) space. The weights $\mu = \varepsilon\tilde{\lambda}/\mathbb{1}^T\tilde{\lambda}$ converge to points at infinity ($\mathbb{1}^T\mu = 0$) of the affine k -plane of normalized weights.

When reference points coalesce with an additional second-order acceleration orthogonally to the subspace W_x , we conjecture that the affine span is not any more a great subspheres but a smaller one. This would include principal nested spheres (PNS) developed by Jung et al. (2010) and Jung, Dryden and Marron (2012) as a limit case of barycentric subspaces. It would be interesting to derive a similar procedure for hyperbolic spaces and to determine which types of subspaces could be obtained by such limits for more general nonlocal and higher-order jets.

6. Barycentric subspace analysis. PCA can be viewed as the search for a sequence of nested linear spaces that best approximate the data at each level. In a Euclidean space, minimizing the variance of the residuals boils down to an independent optimization of orthogonal subspaces at each level of approximation, thanks to the Pythagorean theorem. This enables building 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 the orthogonality of subspaces is not even well defined.

6.1. Flags of barycentric subspaces in manifolds. Damon and Marron (2013) have argued that the nestedness of approximation spaces is one of the most important characteristics for generalizing PCA to more general spaces. Barycentric subspaces can easily be nested, for instance by adding or removing one or several points at a time, to obtain a family of embedded submanifolds which generalizes flags of vector spaces.

A flag of a vector space V is a filtration of subspaces (an increasing sequence of subspaces, where each subspace is a proper subspace of the next): $\{0\} = V_0 \subset V_1 \subset V_2 \subset \dots \subset V_k = V$. Denoting $d_i = \dim(V_i)$ the dimension of the subspaces, we have $0 = d_0 < d_1 < d_2 < \dots < d_k = n$, where n is the dimension of V . Hence, we must have $k \leq n$. A flag is *complete* if $d_i = i$, otherwise it is a *partial flag*. Notice that a linear subspace W of V is identified to the partial flag $\{0\} \subset W \subset V$. A flag can be generated by adding the successive eigenspaces of an SPD matrix with increasing eigenvalues. If all the eigenvalues have multiplicity one, the generated flag is complete and one can parametrize it by the ordered set of eigenvectors. If an eigenvalue has a larger multiplicity, then the corresponding eigenvectors might be considered as exchangeable in this parametrization in the sense that we should only consider the subspace generated by all the eigenvectors of that eigenvalue.

In an n -dimensional manifold \mathcal{M} , a strict ordering of $n + 1$ independent points $x_0 \prec x_1 \prec \dots \prec x_n$ defines a filtration of barycentric subspaces. For instance: $\text{EBS}(x_0) = \{x_0\} \subset \dots \subset \text{EBS}(x_0, x_1, x_k) \subset \dots \subset \text{EBS}(x_0, \dots, x_n)$. The 0-dimensional subspace is now a point in \mathcal{M} instead of the null vector in flags of vector spaces because we are in an affine setting. 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. Indeed, when the manifold $\mathcal{M}^*(x_0, \dots, x_k)$ is connected, the EBS of $n + 1$ affinely independent points covers the full manifold $\mathcal{M}^*(x_0, \dots, x_k)$, and its completion covers the original manifold: $\text{Aff}(x_0, \dots, x_n) = \mathcal{M}$. With the Fréchet or Karcher barycentric subspaces, we only generate a submanifold (the positive span) that does not cover the whole manifold in general, even in negatively curved spaces.

DEFINITION 11 (Flags of affine spans in manifolds). Let $x_0 \preceq x_1 \preceq \dots \preceq x_k$ be $k + 1 \leq n + 1$ affinely independent-ordered points of \mathcal{M} where two or more successive points are either strictly ordered ($x_i \prec x_{i+1}$) or exchangeable ($x_i \sim x_{i+1}$). For a strictly ordered set of points, we call the sequence of properly nested subspaces $\text{FL}_i(x_0 \prec x_1 \prec \dots \prec x_k) = \text{Aff}(x_0, \dots, x_i)$ for $0 \leq i \leq k$ the flag of affine spans $\text{FL}(x_0 \prec x_1 \prec \dots \prec x_k)$. For a flag comprising exchangeable points, the different subspaces of the sequence are only generated at strict ordering signs or at the end. A flag is said complete if it is strictly ordered with $k = n$. We call a flag of exchangeable points $\text{FL}(x_0 \sim x_1 \sim \dots \sim x_k)$ a pure subspace because the sequence is reduced to the unique subspace $\text{FL}_k(x_0 \sim x_1 \sim \dots \sim x_k) = \text{Aff}(x_0, \dots, x_k)$.

6.2. Forward and backward barycentric subspaces analysis. In Euclidean PCA, the flag of linear subspaces can be built in a forward way, by computing the best 0th-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 residual 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 keeping the previous ones fixed. The barycentric subspace $\text{Aff}(x_0) = \{x_0\}$ minimizing the unexplained variance is a Karcher mean. Adding a second point amounts to compute the geodesic passing through the mean that best approximate the data. Adding a third point now differ from PGA, unless the three points coalesce to a single one. With this procedure, the Fréchet mean always belong to the barycentric subspace.

The backward analysis consists in iteratively removing one dimension. One should theoretically start with a full set of points and chose which one to remove. However, as all the sets of $n + 1$ affinely independent points generate the full manifold with the affine span, the optimization really begin with the set of n points

x_0, \dots, x_{n-1} . We should afterward only test for which of the n points we should remove. Since optimization is particularly inefficient in large dimensional spaces, we may run a forward analysis until we reach the noise level of the data for a dimension $k \ll n$. In practice, the noise level is often unknown and a threshold at 5% of the data variance is sometimes chosen. More elaborate methods exist to determine the intrinsic dimension of the data for manifold learning technique [Wang and Marron (2008)]. Point positions may be optimized at each step to find the optimal subspace and a backward sweep reorders the points at the end. With this process, there is no reason for the Fréchet mean to belong to any of the barycentric subspaces. For instance, if we have clusters, one expects the reference points to localize within these clusters rather than at the Fréchet mean.

6.3. *Approximating data using a pure subspace.* Let $Y = \{\hat{y}_i\}_{i=1}^N \in \mathcal{M}^N$ be N data points and $X = \{x_0, \dots, x_k\}$ be $k + 1$ affinely independent reference points. We assume that each data point \hat{y}_i has almost surely one unique closest point $y_i(X)$ on the barycentric subspace. This is the situation for Euclidean, hyperbolic and spherical spaces, and this should hold more generally for all the points outside the focal set of the barycentric subspace. This allows us to write the residual $r_i(X) = \text{dist}(\hat{y}_i, y_i(X))$ and to consider the minimization of the unexplained variance $\sigma_{\text{out}}^2(X) = \sum_j r_j^2(X)$. This optimization problem on \mathcal{M}^{k+1} can be achieved by standard techniques of optimization on manifolds [see, e.g., Absil, Mahony and Sepulchre (2008)]. However, it is not obvious that the canonical product Riemannian metric is the right metric to use, especially close to coincident points. In this case, one would like to consider switching to the space of (nonlocal) jets to guaranty the numerical stability of the solution. In practice, though, we may constraint the distance between reference points to be larger than a threshold.

A second potential problem is the lack of identifiability: the minimum of the unexplained variance may be reached by subspaces parametrized by several k -tuples of points. This is the case for constant curvature spaces since every linearly independent k -tuple of points in a given subspace parametrizes the same barycentric subspace. In constant curvature spaces, this can be accounted for using a suitable polar or QR matrix factorization [see, e.g., Pennec (2017b)]. In general manifolds, we expect that the absence of symmetries will break the multiplicity of this relationship (at least locally) thanks to the curvature. However, it can lead to very badly conditioned systems to solve from a numerical point of view for small curvatures.

A last problem is that the criterion we use here (the unexplained variance) is only valid for a pure subspace of fixed dimension, and considering a different dimension will lead in general to pure subspaces which cannot be described by a common subset of reference points. Thus, the forward and backward optimization of nested barycentric subspaces cannot lead to the simultaneous optimality of all the subspaces of a flag in general manifolds.

6.4. *A criterion for hierarchies of subspaces: AUV on flags of affine spans.* In order to obtain consistency across dimensions, it is necessary to define a criterion which depends on the whole flag of subspaces and not on each of the subspaces independently. In PCA, one often plots the unexplained variance as a function of the number of modes used to approximate the data. This curve should decrease as fast as possible from the variance of the data (for 0 modes) to 0 (for n modes). A standard way to quantify the decrease consists in summing the values at all steps, giving the Accumulated Unexplained Variances (AUV), which is analogous to the Area-Under-the-Curve (AUC) in Receiver Operating Characteristic (ROC) curves.

Given a strictly ordered flag of affine subspaces $\text{Fl}(x_0 \prec x_1 \prec \dots \prec x_k)$, we thus propose to optimize the AUV criterion:

$$\text{AUV}(\text{Fl}(x_0 \prec x_1 \prec \dots \prec x_k)) = \sum_{i=0}^k \sigma_{\text{out}}^2(\text{Fl}_i(x_0 \prec x_1 \prec \dots \prec x_k))$$

instead of the unexplained variance at order k . We could of course consider a complete flag but in practice it is often useful to stop at a dimension k much smaller than the possibly very high dimension n . The criterion is extended to more general partial flags by weighting the unexplained variance of each subspace by the number of (exchangeable) points that are added at each step. With this global criterion, the point x_i influences all the subspaces of the flag that are larger than $\text{Fl}_i(x_0 \prec x_1 \prec \dots \prec x_k)$ but not the smaller subspaces. It turns out that optimizing this criterion results in the usual PCA up to mode k in a Euclidean space.

THEOREM 9 (Euclidean PCA as an optimization in the flag space). *Let $\hat{Y} = \{\hat{y}_i\}_{i=1}^N$ be a set of N data points in \mathbb{R}^n . We denote as usual the mean by $\bar{y} = \frac{1}{N} \sum_{i=1}^N \hat{y}_i$ and the empirical covariance matrix by $\Sigma = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - \bar{y})(\hat{y}_i - \bar{y})^T$. Its spectral decomposition is denoted by $\Sigma = \sum_{j=1}^n \sigma_j^2 u_j u_j^T$ with the eigenvalues sorted in decreasing order. We assume that the first $k + 1$ eigenvalues have multiplicity one, so that the order from σ_1 to σ_{k+1} is strict.*

Then the partial flag of affine subspaces $\text{Fl}(x_0 \prec x_1 \prec \dots \prec x_k)$ optimizing

$$\text{AUV}(\text{Fl}(x_0 \prec x_1 \prec \dots \prec x_k)) = \sum_{i=0}^k \sigma_{\text{out}}^2(\text{Fl}_i(x_0 \prec x_1 \prec \dots \prec x_k))$$

is strictly ordered and can be parametrized by $x_0 = \bar{y}$, $x_i = x_0 + u_i$ for $1 \leq i \leq k$. The parametrization by points is not unique but the flag of subspaces which is generated is and is equal to the flag generated by the PCA modes up to mode k included.

The proof is detailed in [Pennec \(2017b\)](#). The main idea is to parametrize the matrix of reference vectors by the product of an orthogonal matrix Q with a positive definite triangular superior matrix (QR decomposition). The key property of this

Gram–Schmidt orthogonalization is the stability of the columns of Q when we add or remove columns (i.e., reference points) in X , which allows to write the expression of the AUV explicitly. Critical points are found for columns of Q which are eigenvectors of the data covariance matrix and the expression of the AUV shows that we have to select them in the decreasing order of eigenvalues.

6.5. *Sample-limited barycentric subspace inference on spheres.* In several domains, it has been proposed to limit the inference of the Fréchet mean to the data-points only. In neuroimaging studies, for instance, the individual image minimizing the sum of square deformation distance to other subject images has been argued to be a good alternative to the mean template (a Fréchet mean in deformation and intensity space) because it conserves the full definition and all the original characteristics of a real subject image [Leporé et al. (2008)]. Beyond the Fréchet mean, Feragen et al. (2013) proposed to define the first principal component mode as the geodesic going through two of the data points which minimizes the unexplained variance. The method named *set statistics* was aiming to accelerate the computation of statistics on tree spaces. Zhai (2016) further explored this idea under the name of *sample-limited geodesics* in the context of PCA in phylogenetic tree space. However, in both cases, extending the method to higher-order principal modes was considered as a challenging research topic.

With barycentric subspaces, sample-limited statistics naturally extends to any dimension by restricting the search to (flags of) affine spans that are parametrized by data points. Moreover, the implementation boils down to a very simple enumeration problem. An important advantage for interpreting the modes of variation is that reference points are never interpolated as they are by definition sampled from the data. Thus, we may go back to additional information about the samples like the disease characteristics in medical image analysis. The main drawback is the combinatorial explosion of the computational complexity: the optimal order- k flag of affine spans requires $O(N^{k+1})$ operations, where N is the number of data points. In practice, the search can be done exhaustively for a small number of reference points but an approximated optimum has to be sought for larger k using a limited number of random tuples [Feragen et al. (2013)].

In this section, we consider the exhaustive sample-limited version of the Forward Barycentric Subspace (FBS) decomposition, the optimal k -dimensional Pure Barycentric Subspace with backward ordering (k -PBS), and the Barycentric Subspace Analysis up to order k (k -BSA). In order to illustrate the differences, we consider a first synthetic dataset where we draw 30 random points uniformly on an equilateral triangle of side length $\pi/2$ on a 6-dimensional sphere. We add to each point a (wrapped) Gaussian noise of standard deviation $\sigma = 10^\circ$. In this example, original data live on a 2-sphere: the ideal flag of subspaces is a pure 2d subspace spanning the first three coordinates. We illustrate in Figure 4 the different reference points that are found for the different methods. We can see that all methods end-up with different results, contrarily to the Euclidean case. The second

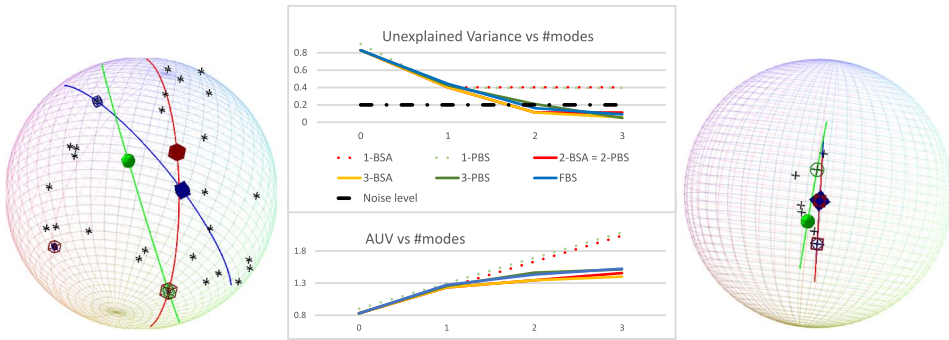


FIG. 4. Left: *Equi 30* simulated dataset. Data and reference points are projected from the 5-sphere to the expected 2-sphere in 3d to allow visualization. For each method (FBS in blue, 1-PBS in green and 1-BSA in red), the first reference point has a solid symbol. The 1d mode is the geodesic joining this point to the second reference point. The third reference point of FBS and 2-BSA (on the lower left part) is smaller. Middle: graph of the unexplained variance and AUV for the different methods on the *Equi 30* dataset. Right: *Mount Tom Dinosaur trackway 1* data with the same color code. 1-BSA (in red) and FBS (in blue) are superimposed.

observation is that the optimal pure subspace is not stable with the dimension: the reference points of the 0-PBS (the sample-limited Fréchet mean represented by the large blue solid diamond), the 1-PBS (in green) and the 2-PBS (identical to the red points of the 2-BSA in red) are all different. BSA is more stable: the first reference points are the same from the 1-BSA to the 3-BSA. In terms of unexplained variance, the 2-BSA is the best for two modes (since it is identical to the optimal 2-PBS) and reaches the actual noise level. It remains better than the 3-PBS and the FBS with three modes in terms of AUV even without adding a fourth point.

As a second example, we take real data encoding the shape of three successive footprints of Mount Tom Dinosaur trackway 1 described in Small (1996), page 181. For planar triangles, the shape space (quotient of the triad by similarities) boils down to the sphere of radius $1/2$. These data are displayed on the right of Figure 4. In this example, the reference points of the 0-BSA to the 3-BSA are stable and identical to the ones of the FBS. This is a behavior that we have observed in most of our simulations when modes cannot be confused. This may not hold anymore if reference points were optimized on the sphere rather than on the data points only. The optimal 1-PBS (the best geodesic approximation) picks up different reference points.

7. Discussion. We investigated in the paper several notions of subspaces in manifolds generalizing the notion of affine span in a Euclidean space. The Fréchet/Karcher/exponential barycentric subspaces are the nested locus of weighted Fréchet/Karcher/exponential barycenters with positive or negative weights summing up to 1. The affine spans is the metric completion of the largest

one (the EBS). It may be a nonconnected manifold with boundaries. The completeness of the affine span enables reconnecting part of the subspace that arrive from different directions at the cut-locus of reference points if needed. It also ensures that there exists a closest point on the submanifold for data projection purposes, which is fundamental for dimension reduction purposes. The fact that modifying the power of the metric does not change the affine span is an unexpected stability result which suggests that the notion is quite central. Moreover, we have shown that the affine span encompasses principal geodesic subspaces as limit cases. It would be interesting to show that we can obtain other types of subspaces like principal nested subspheres with higher-order and nonlocal jets: some nongeodesic decomposition schemes such as loxodromes and splines could probably also be seen as limit cases of barycentric subspaces.

Future work will address barycentric subspaces in interesting nonconstant curvatures spaces. For instance, [Eltzner, Jung and Huckemann \(2015\)](#) adaptively deforms the flat torus seen as a product of spheres into a unique sphere to allow principal nested spheres (PNS) analysis. A quick look at the flat torus shows that the the cut-locus of $k + 1 \leq n$ points in \mathcal{S}_1^n divides the torus into k^n cells in which the affine span is a k -dimensional linear subspace. The subspaces generated in each cell are generally disconnected, but when points coalesce with each others into a jet, the volume of all but one cells decrease and at the limit we recover a single non-empty cell that contain a connected affine span. For a first-order jet, we recover as expected the restricted geodesic subspace (here a linear subspace limited to the cut locus of the jet base-point), but higher-order jets may generate more interesting curved subspaces that may better describe the data geometry.

The next practical step is obviously the implementation of generic algorithms to optimize barycentric subspaces in general Riemannian manifolds. Example algorithms include: finding a point with given barycentric coordinates (there might be several so this has to be a local search); finding the closest point (and its coordinates) on the barycentric subspace; optimizing the reference points to minimize the residual error after projection of data points, etc. If such algorithms can be designed relatively simply for simple specific manifolds as we have done here for constant curvature spaces, the generalization to general manifolds requires a study of the focal set of the barycentric subspaces or guarantying the correct behavior of algorithms. We conjecture that this is a stratified set of zero measure in generic cases. Another difficulty is linked to the nonidentifiability of the subspace parameters. For constant curvature spaces, the right parameter space is actually the k -Grassmanian. In more general manifolds, the curvature and the interaction with the cut-locus break the symmetry of the barycentric subspaces, but lead to a poor numerical conditioning of the system good renormalization techniques need to be designed to guaranty the numerical stability.

Finding the subspace that best explain the data is an optimization problem on manifolds. This raises the question of which metric should be considered on the

space of barycentric subspaces. In this paper, we mainly see this space as the configuration space of $k + 1$ affinely independent points, with convergence to spaces of jets (including nonlocal jets) when several points coalesce. Such a construction was named Multispace by [Olver \(2001\)](#) in the context of symmetry-preserving numerical approximations to differential invariants. It is likely that similar techniques could be investigated to construct numerically stable implementations of barycentric subspaces of higher-order parametrized by nonlocal jets, which are needed to optimize safely. Conversely, barycentric subspaces could help shedding a new light on the multispace construction for differential invariants.

Barycentric subspaces could probably be used to extend methods like the probabilistic PCA of [Tipping and Bishop \(1999\)](#), generalized to PGA by [Zhang and Fletcher \(2013\)](#). A first easy step in that direction is to replace the reference points by reference distributions on the manifold and to look at the locus of weighted expected means. Interestingly, this procedure softens the constraints that we had in this paper about the cut locus. Thus, following [Karcher \(1977\)](#), reference distributions could be used in a mollifier smoothing approach to study the regularity of the barycentric subspaces.

For applications where data live on Lie groups, generalizing barycentric subspaces to more general non-Riemannian spaces like affine connection manifolds is a particularly appealing extension. In computational anatomy, for instance, deformations of shapes are lifted to a group of diffeomorphism for statistical purposes [see, e.g., [Lorenzi and Pennec \(2013\)](#) and [Lorenzi, Ayache and Pennec \(2015\)](#)]. All Lie groups can be endowed with a bi-invariant symmetric Cartan–Schouten connection for which geodesics are the left and right translation of one-parameter subgroups. This provides the Lie group with an affine connection structure which may be metric or not. When the group is the direct product of compact and Abelian groups, it admits a bi-invariant metric for which the Cartan–Schouten connection is the natural Levi–Civita connection. Other groups do not admit any bi-invariant metric (this is the case for rigid transformations in more than 2 dimensions because of the semi-direct product), so that a Riemannian structure can only be left or right invariant but not both. However the bi-invariant Cartan–Schouten connection continues to exist, and one can design bi-invariant means using exponential barycenter as proposed by [Pennec and Arsigny \(2013\)](#). Thus, we may still define exponential barycentric subspaces and affine spans in these affine connection spaces, the main difference being that the derivative of the log is not any more the Hessian of a distance function. This might considerably complexify the analysis of the generated subspaces.

The second topic of this paper concerns the generalization of PCA to manifolds using Barycentric Subspace Analysis (BSA). [Damon and Marron \(2013\)](#) argued that an interesting generalization of PCA should rely on “nested sequence of relations”, like embedded linear subspaces in the Euclidean space or embedded spheres in PNS. Barycentric subspaces can naturally be nested by adding or removing points or equivalently by setting the corresponding barycentric coordinate

to zero. Thus, we can easily generalize PCA to manifolds using a forward analysis by iteratively adding one or more points at a time. At the limit where points coalesce at the first order, this amounts to build a flag of (restricted) principal geodesic subspaces. Thus, it generalizes the Principal Geodesic Analysis (PGA) of Fletcher et al. (2004) and Sommer, Lauze and Nielsen (2014) when starting with a zeroth dimensional space (the Fréchet mean) and the Geodesic PCA (GPCA) of Huckemann and Ziezold (2006) and Huckemann, Hotz and Munk (2010) when starting directly with a first-order jet defining a geodesic. One can also design a backward analysis by starting with a large subspace and iteratively removing one or more points to define embedded subspaces.

However, the greedy optimization of these forward/backward methods generally leads to different solutions which are not optimal for all subspace jointly. The key idea is to consider PCA as a joint optimization of the whole flag of subspaces instead of each subspace independently. In a Euclidean space, we showed that the Accumulated Unexplained Variances (AUV) with respect to all the subspaces of the hierarchy (the area under the curve of unexplained variance) is a proper criterion on the space of Euclidean flags. We proposed to extend this criterion to barycentric subspaces in manifolds, where an ordering of the reference points naturally defines a flag of nested barycentric subspaces. A similar idea could be used with other iterative least-squares methods like partial least-squares (PLS) which are also one-step at a time minimization methods.

APPENDIX: PROOF OF THEOREM 8

PROOF OF THEOREM 8. We first establish a useful formula exploiting the symmetry of the geodesics from x to $y \notin \mathcal{C}(x)$ with respect to time. Reverting time along a geodesic, we have: $\gamma_{(x, \vec{x}\vec{y})}(t) = \gamma_{(y, \vec{y}\vec{x})}(1 - t)$, which means in particular that $\dot{\gamma}_{(x, \vec{x}\vec{y})}(1) = -\dot{\gamma}_{(y, \vec{y}\vec{x})}(0) = -\vec{y}\vec{x}$. Since $\gamma_{(x, \vec{x}\vec{y})}(t) = \exp_x(t\vec{x}\vec{y})$, we obtain $\vec{y}\vec{x} = -D \exp_x |_{\vec{x}\vec{y}} \vec{x}\vec{y}$. Now, we also have $(D \exp_x |_{\vec{x}\vec{y}}) \cdot D \log_x |_y = \text{Id}$ because $\exp_x(\log_x(y)) = y$. Finally, $D \exp_x$ and $D \log_x$ have full rank on $\mathcal{M}/\mathcal{C}(x)$ since there is no conjugate point before the cut-locus, so that we can multiply by their inverse and we end up with

$$(16) \quad \forall y \notin \mathcal{C}(x), \quad \vec{x}\vec{y} = -D \log_x |_y \vec{y}\vec{x}.$$

Let us first restrict to a convenient domain of \mathcal{M} : we consider a open geodesic ball $B(x_0, \zeta)$ of radius ζ centered at x_0 and we exclude all the points of \mathcal{M} which cut locus intersect this ball, or equivalently the cut-locus of all the points of this ball. We obtain an open domain $\mathcal{D}_\zeta(x_0) = \mathcal{M} \setminus \mathcal{C}(B(x_0, \zeta))$ in which $\log_x(y)$ is well defined and smooth for all $x \in B(x_0, \zeta)$ and all $y \in \mathcal{D}_\zeta(x_0)$. Thanks to the symmetry of the cut-locus, $\log_y(x)$ is also well defined and smooth in the same conditions and equation (16) can be rephrased:

$$(17) \quad \forall x \in B(x_0, \zeta), y \in \mathcal{D}_\zeta(x_0), \quad \vec{x}\vec{y} = -D \log_x |_y \vec{y}\vec{x}.$$

Let $\|w\|_\infty = \max_i \|w_i\|_{x_0}$ be the maximal length of the vectors w_i . For $\varepsilon < \zeta / \|w\|_\infty$, we have $\|\varepsilon w_i\|_{x_0} \leq \varepsilon \|w\|_\infty < \zeta$, so that all the points $x_i = \exp_{x_0}(\varepsilon w_i)$ belong to the open geodesic ball $B(x_0, \zeta)$. Thus, $\log_x(x_i)$ and $\log_{x_i}(x)$ are well defined and smooth for any $x \in \mathcal{D}_\zeta(x_0)$, and we can write the Taylor expansion in a normal coordinate system at x_0 using equation (17):

$$\begin{aligned} \log_x(x_i(\varepsilon)) &= \log_x(x_0) + \varepsilon D \log_x|_{x_0} w_i + O(\varepsilon^2) \\ &= D \log_x|_{x_0}(\varepsilon w_i - \log_{x_0}(x)) + O(\varepsilon^2). \end{aligned}$$

Any point $x \in \mathcal{D}_\zeta(x_0)$ can be defined by $\log_{x_0}(x) = \sum_{j=1}^k \alpha_j w_j + w_\perp$ with $\langle w_\perp | w_j \rangle = 0$ and suitable constraints on the α_i and w_\perp . Replacing $\log_x(x_0)$ by its value in the above formula, we get

$$\log_x(x_i(\varepsilon)) = D \log_x|_{x_0} \left(\varepsilon w_i - \sum_{j=1}^k \alpha_j w_j - w_\perp \right) + O(\varepsilon^2).$$

Since the matrix $D \log_x|_{x_0}$ is invertible, the EBS equation $\mathfrak{M}_1(x, \lambda) = \sum_{i=0}^k \lambda_i \vec{x}_i = 0$ is equivalent to $w_\perp + \sum_{j=1}^k \alpha_j w_j - \varepsilon (\sum_{i=1}^k \lambda_i w_i) = O(\varepsilon^2)$. Projecting orthogonally to W_{x_0} , we get $w_\perp = O(\varepsilon^2)$: this means that any point of the limit EBS has to be of the form $x = \exp_{x_0}(\sum_{j=1}^k \alpha_j w_j)$. In other words, only points of the restricted geodesic subspace $GS^*(W_{x_0})$ can be solutions of the limit EBS equation.

Now, for a point of $GS^*(W_{x_0})$ to be a solution of the limit EBS equation, there should exist barycentric coordinates λ such that $\sum_{j=1}^k (\alpha_j - \varepsilon \lambda_j) w_j = O(\varepsilon^2)$. Choosing $\lambda = (\varepsilon - \sum_i \alpha_i : \alpha_1 : \dots : \alpha_k)$, we obtain the normalized barycentric coordinates $\lambda_j = \alpha_j / \varepsilon$ for $1 \leq j \leq k$ and $\lambda_0 = 1 - (\sum_i \alpha_i) / \varepsilon$ that satisfy this condition. Thus, any point of $GS^*(W_{x_0}) \cap \mathcal{D}_\zeta(x_0)$ is a solution of the limit EBS equation with barycentric coordinates at infinity on \mathcal{P}_k^* . Taking ζ sufficiently small, we can include all the points of $GS^*(W_{x_0})$. \square

Acknowledgments. I would particularly like to thank Professor Susan Holmes for fruitful discussions during the writing of the paper.

SUPPLEMENTARY MATERIAL

Supplement A: Hessian of the Riemannian squared distance (DOI: [10.1214/17-AOS1636SUPPA](https://doi.org/10.1214/17-AOS1636SUPPA); .pdf). This supplementary material describes in more length the notions of Riemannian geometry that are underlying the main paper and investigates the Hessian of the Riemannian square distance whose eigenvalues control the local regularity of the barycentric subspaces. This is exemplified on the sphere and the hyperbolic space.

Supplement B: PCA as an optimization on the flag manifold (DOI: [10.1214/17-AOS1636SUPPB](https://doi.org/10.1214/17-AOS1636SUPPB); .pdf). This supplementary material details in length the proof that the flag of linear subspaces found by PCA optimizes the Accumulated Unexplained Variances (AUV) criterion in a Euclidean space.

REFERENCES

- ABSIL, P.-A., MAHONY, R. and SEPULCHRE, R. (2008). *Optimization Algorithms on Matrix Manifolds*. [MR2364186](#)
- AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. [MR2736346](#)
- BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. [MR1962498](#)
- BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. [MR2195634](#)
- BISHOP, R. L. and O’NEILL, B. (1969). Manifolds of negative curvature. *Trans. Amer. Math. Soc.* **145** 1–49. [MR251664](#)
- BREWIN, L. (2009). Riemann normal coordinate expansions using Cadabra. *Classical Quantum Gravity* **26** 175017. [MR2534336](#)
- BUSEMANN, H. (1955). *The Geometry of Geodesics*. Academic Press, San Diego. [MR0075623](#)
- BUSER, P. and KARCHER, H. (1981). *Gromov’s Almost Flat Manifolds*. *Astérisque* **81**. Société Mathématique de France, Paris. [MR0619537](#)
- BUSS, S. R. and FILLMORE, J. P. (2001). Spherical averages and applications to spherical splines and interpolation. *ACM Trans. Graph.* **20** 95–126.
- COSTA, S. I. R., SANTOS, S. A. and STRAPASSON, J. E. (2015). Fisher information distance: A geometrical reading. *Discrete Appl. Math.* **197** 59–69. [MR3398961](#)
- CROUCH, P. and LEITE, F. S. (1995). The dynamic interpolation problem: On Riemannian manifolds, Lie groups, and symmetric spaces. *J. Dyn. Control Syst.* **1** 177–202. [MR1333770](#)
- CUTLER, A. and BREIMAN, L. (1994). Archetypal analysis. *Technometrics* **36** 338–347. [MR1304898](#)
- DAMON, J. and MARRON, J. S. (2013). Backwards principal component analysis and principal nested relations. *J. Math. Imaging Vision* **50** 107–114. [MR3233137](#)
- DRYDEN, I. L. (2005). Statistical analysis on high-dimensional spheres and shape spaces. *Ann. Statist.* **33** 1643–1665. [MR2166558](#)
- ELTZNER, B., JUNG, S. and HUCKEMANN, S. (2015). Dimension reduction on polyspheres with application to skeletal representations. In *Geometric Science of Information* (F. Nielsen and F. Barbaresco, eds.). *Lecture Notes in Computer Science* **9389** 22–29. Springer, Cham. [MR3442181](#)
- FERAGEN, A., OWEN, M., PETERSEN, J., WILLE, M. M. W., THOMSEN, L. H., DIRKSEN, A. and DE BRUIJNE, M. (2013). Tree-space statistics and approximations for large-scale analysis of anatomical trees. In *International Conference on Information Processing in Medical Imaging (IPMI 2013)*, Asilomar, CA, USA. *Lecture Notes in Computer Science* **7917** 74–85. Springer, Berlin.
- FLETCHER, P. T., LU, C., PIZER, S. M. and JOSHI, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imag.* **23** 995–1005.
- FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré* **10** 215–310. [MR27464](#)
- GAY-BALMAZ, F., HOLM, D., MEIER, D., RATIU, T. and VIALARD, F.-X. (2012). Invariant higher-order variational problems. *Comm. Math. Phys.* **309** 413–458. [MR2864799](#)
- GROSSER, D. (2004). Newton’s method, zeroes of vector fields, and the Riemannian center of mass. *Adv. in Appl. Math.* **33** 95–135. [MR2064359](#)
- HUCKEMANN, S., HOTZ, T. and MUNK, A. (2010). Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statist. Sinica* **20** 1–58. [MR2640651](#)
- HUCKEMANN, S. and ZIEZOLD, H. (2006). Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces. *Adv. in Appl. Probab.* **38** 299–319. [MR2264946](#)

- JUNG, S., DRYDEN, I. L. and MARRON, J. S. (2012). Analysis of principal nested spheres. *Biometrika* **99** 551–568. [MR2966769](#)
- JUNG, S., LIU, X., MARRON, J. S. and PIZER, S. M. (2010). Generalized PCA via the backward stepwise approach in image analysis. In *Proc. of the Int. Symposium Brain, Body and Machine. Advances in Intelligent and Soft Computing* **83** 111–123. Springer, Berlin.
- KARCHER, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30** 509–541. [MR442975](#)
- KARCHER, H. (2014). Riemannian center of mass and so called Karcher mean. Available at [arXiv:1407.2087](#).
- KENDALL, W. S. (1990). Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence. *Proc. Lond. Math. Soc.* (3) **61** 371–406. [MR1063050](#)
- LE, H. (2004). Estimation of Riemannian barycentres. *LMS J. Comput. Math.* **7** 193–200. [MR2085875](#)
- LEPORÉ, N., BRUN, C., CHOU, Y.-Y., LEE, A., BARYSHEVA, M., PENNEC, X., MCMAHON, K., MEREDITH, M., DE ZUBICARAY, G., WRIGHT, M., TOGA ARTHUR, W. and THOMPSON, P. (2008). Best individual template selection from deformation tensor minimization. In *Proc. of the 2008 IEEE Int. Symp. ISBI 2008, Paris, France* 460–463.
- LORENZI, M., AYACHE, N. and PENNEC, X. (2015). Regional flux analysis for discovering and quantifying anatomical changes: An application to the brain morphometry in Alzheimer’s disease. *NeuroImage* **115** 224–234.
- LORENZI, M. and PENNEC, X. (2013). Geodesics, parallel transport & one-parameter subgroups for diffeomorphic image registration. *Int. J. Comput. Vis.* **105** 111–127. [MR3104013](#)
- MACHADO, L., SILVA LEITE, F. and KRAKOWSKI, K. (2010). Higher-order smoothing splines versus least squares problems on Riemannian manifolds. *J. Dyn. Control Syst.* **16** 121–148. [MR2580471](#)
- OLVER, P. J. (2001). Geometric foundations of numerical algorithms and symmetry. *Appl. Algebra Engrg. Comm. Comput.* **11** 417–436. [MR1828189](#)
- PENNEC, X. (2006). Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vision* **25** 127–154. [MR2254442](#)
- PENNEC, X. (2015). Barycentric subspaces and affine spans in manifolds. In *Geometric Science of Information. Lecture Notes in Computer Science* **9389** 12–21. Springer, Cham. [MR3442180](#)
- PENNEC, X. (2018a). Supplement A to “Barycentric subspace analysis on manifolds.” DOI:[10.1214/17-AOS1636SUPPA](#).
- PENNEC, X. (2018b). Supplement B to “Barycentric subspace analysis on manifolds.” DOI:[10.1214/17-AOS1636SUPPB](#).
- PENNEC, X. and ARSIGNY, V. (2013). Exponential barycenters of the canonical Cartan connection and invariant means on Lie groups. In *Matrix information geometry* 123–166. Springer, Heidelberg. [MR2964451](#)
- PENNEC, X., FILLARD, P. and AYACHE, N. (2006). A Riemannian framework for tensor computing. *Int. J. Comput. Vis.* **66** 41–66.
- SMALL, C. G. (1996). *The Statistical Theory of Shapes*. Springer, Berlin. [MR1418639](#)
- SOMMER, S. (2013). Horizontal dimensionality reduction and iterated frame bundle development. In *Proceedings of the 1st International Conference Geometric Science of Information Held in Paris, August 28–30, 2013 (GSI 2013)* (F. Nielsen and F. Barbaresco, eds.). *Lecture Notes in Computer Science* **8085** 76–83. Springer, Heidelberg. [MR3126126](#)
- SOMMER, S., LAUZE, F. and NIELSEN, M. (2014). Optimization over geodesics for exact principal geodesic analysis. *Adv. Comput. Math.* **40** 283–313. [MR3194707](#)
- TIPPING, M. E. and BISHOP, C. M. (1999). Probabilistic principal component analysis. *J. Roy. Statist. Soc. Ser. B* **61** 611–622. [MR1707864](#)
- WANG, X. and MARRON, J. S. (2008). A scale-based approach to finding effective dimensionality in manifold learning. *Electron. J. Stat.* **2** 127–148. [MR2386090](#)

- WEYENBERG, G. S. (2015). Statistics in the Billera–Holmes–Vogtmann treespace. Ph.D. thesis, Univ. Kentucky. [MR3450194](#)
- WILSON, R. C., HANCOCK, E. R., PEKALSKA, E. and DUIN, R. P. W. (2014). Spherical and hyperbolic embeddings of data. *IEEE Trans. Pattern Anal. Mach. Intell.* **36** 2255–2269.
- YANG, L. (2011). Medians of probability measures in Riemannian manifolds and applications to radar target detection. Ph.D. thesis, Poitiers Univ.
- ZHAI, H. (2016). Principal component analysis in phylogenetic tree space. Ph.D. thesis, Univ. North Carolina at Chapel Hill, Ann Arbor, MI. [MR3542232](#)
- ZHANG, M. and FLETCHER, P. T. (2013). Probabilistic principal geodesic analysis. In *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'13)* **1** 1178–1186.

UNIVERSITÉ CÔTE D'AZUR
INRIA, SOPHIA ANTIPOLIS MÉDITERRANÉE
ASCLEPIOS TEAM
2004 ROUTE DES LUCIOLES BP 93
F-06902 SOPHIA ANTIPOLIS CEDEX
FRANCE
E-MAIL: xavier.pennec@inria.fr