



HAL
open science

Advances in Geometric Statistics for Manifold Dimension Reduction

Xavier Pennec

► **To cite this version:**

Xavier Pennec. Advances in Geometric Statistics for Manifold Dimension Reduction. Handbook of Variational Methods for Nonlinear Geometric Data, Springer International Publishing, pp.339-359, 2020, 10.1007/978-3-030-31351-7_11 . hal-02560938

HAL Id: hal-02560938

<https://inria.hal.science/hal-02560938>

Submitted on 2 May 2020

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

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

Advances in Geometric Statistics for manifold dimension reduction*

Xavier Pennec

Abstract

Geometric statistics aim at shifting the classical paradigm for inference from points in a Euclidean space to objects living in a non-linear space, in a consistent way with the underlying geometric structure considered. In this chapter, we illustrate some recent advances of geometric statistics for dimension reduction in manifolds. Beyond the mean value (the best 0-dimensional summary statistics of our data), we want to estimate higher dimensional approximation spaces fitting our data. We first define a family of natural parametric geometric subspaces in manifolds that generalize the now classical geodesic subspaces: barycentric subspaces are implicitly defined as the locus of weighted means of $k + 1$ reference points with positive or negative weights summing up to one. Depending on the definition of the mean, we obtain the Fréchet, Karcher or Exponential Barycentric subspaces (FBS/KBS/EBS). The completion of the EBS, called the affine span of the points in a manifold is the most interesting notion as it defines complete sub-(pseudo)-spheres in constant curvature spaces. Barycentric subspaces can be characterized very similarly to the Euclidean case by the singular value decomposition of a certain matrix or by the diagonalization of the covariance and the Gram matrices. This shows that they are stratified spaces that are locally manifolds of dimension k at regular points.

Barycentric subspaces can naturally be nested by defining an ordered series of reference points in the manifold. This allows the construction of inductive forward or backward properly nested sequences of subspaces approximating data points. These flags of barycentric subspaces generalize the sequence of nested linear subspaces (flags) appearing in the classical Principal Component Analysis. We propose a criterion on the space of flags, the accumulated unexplained variance (AUV), whose optimization exactly lead to the PCA decomposition in Euclidean spaces. This procedure is called barycentric subspace analysis (BSA). We illustrate the power of barycentric subspaces in the context of cardiac imaging with the estimation, analysis and reconstruction of cardiac motion from sequences of images.

1 Introduction

Statistical computing on simple manifolds like spheres or flat tori raises problems due to the non-linearity of the space. For instance, averaging points on a sphere using the properties of the embedding Euclidean space leads to a point located inside the sphere and not on its surface. More generally, the classical mean value of random numeric values with distribution P is defined through an integral $\bar{x} = \int x dP(x)$, which can be rewritten as an implicit barycentric equation $\int (x - \bar{x}) dP(x) = 0$. Notice that this notion is intrinsically affine. However, since an integral is a linear operator, this definition of the mean is bound to fail for general non-linear spaces. Geometric statistics were born out of this observation by Maurice Fréchet in the 1940's.

Geometric statistics is a rapidly evolving domain at the confluent of several mathematical and application domains. It was driven in the 80's by Kendall's shape spaces, which encode the configuration of k points under a transformation group, often rigid body transformations or similarities, see e.g [24, 9, 29, 46]. Applied mathematicians and computer scientists got

*Author pre-print of chapter to appear in *Handbook of Variational Methods for Nonlinear Geometric Data*, P. Grohs et al. (eds.), Springer 2020. https://doi.org/10.1007/978-3-030-31351-7_11

interested in the 90's in computing and optimizing on specific matrix manifolds, like rotations, rigid body transformations, Stiefel and Grassmann manifolds [10, 42, 17, 16, 34]. In the context of computer vision and medical image analysis applications, the Fréchet mean was used to develop a practical toolbox for statistics on manifolds in the 90ies [36, 37], with applications to the uncertainty of medical image registration and statistics on shapes. For instance, statistical distances such as the Mahalanobis distance were developed to define some simple statistical tests on manifolds. With the example of diffusion tensor images, this statistical toolbox was generalized in [41] to many manifold valued image processing algorithms such as interpolation, filtering, diffusion and restoration of missing data. Some of these statistics were generalized to even more non-Euclidean data like trees and graphs with object-oriented data analysis [32].

In this chapter, we illustrate some recent advances for dimension reduction in manifolds. Beyond the mean value, which is a good 0-dimensional summary statistic of our data, we want to estimate higher dimensional approximation spaces fitting our data. In [39], we have proposed a new and general family of subspaces in manifolds, *barycentric subspaces*, implicitly defined as the locus of weighted means of $k + 1$ reference points. Barycentric subspaces can naturally be nested and allow the construction of inductive forward or backward nested subspaces approximating data points. We can also consider the whole hierarchy of embedded barycentric subspaces defined by an ordered series of points in the manifold (a flag of affine spans): optimizing the accumulated unexplained variance (AUV) over all the subspaces actually generalizes PCA to non Euclidean spaces, a procedure named Barycentric Subspaces Analysis (BSA). We illustrate the power of barycentric subspaces in the context of cardiac imaging with the estimation, analysis and reconstruction of cardiac motion from sequences of images.

2 Means on manifolds

The Fréchet mean Maurice Fréchet was the first to try to generalize the notion of mean, median and other types of typical central values to abstract spaces for statistical purposes. In a preparatory work [14] he first investigated different ways to compute mean values of random triangles, independently of their position and orientation in space. In the same paper he reports a series of experiments with careful measurements to reproduce the theoretical values. In this respect, he was really pioneering the statistical study of shapes. In a second study, motivated by the study of random curves, he first introduced a mean value and a law of large numbers defined by a generalization of the integral to normed vector (Wiener or Banach) spaces. Finally, Fréchet considered in [15] the generalization of many central values (including the mean and the median) to random elements in abstract metric spaces.

Definition 1 (Fréchet mean in a metric space) [15, p.233] *The p -mean (typical position of order p according to Fréchet) of a distribution μ (a random element) in an abstract metric space \mathcal{M} is set of minimizers of the mean p -distance (MpD):*

$$Mean_p(\mu) = \left\{ \arg \min_{y \in \mathcal{M}} \frac{1}{p} \int_{\mathcal{M}} \text{dist}(x, y)^p d\mu(x) \right\}. \quad (1)$$

In a Euclidean space, this defines the usual arithmetic mean for $p = 2$, the median (valeur équiprobable in Fréchet's words) for $p = 1$ and the barycenter of the support of the distribution for $p = \infty$.

The first key contribution of Fréchet was to consider many different types of typical elements, including of course the mean but also the median. Fréchet considered mainly the case $p \geq 1$, but he observed that many of the properties could be also generalized to $0 < p < 1$. The second key contribution of Fréchet was to consider the mean as a set of elements rather than

one unique element¹. This key innovation was later developed by Ziezold [54] with a strong law of large numbers for sets of random elements in separable finite quasi-metric spaces. These two innovations justify the name naming of Fréchet mean that is used in geometric statistics.

The Riemannian center of mass In a complete metric space, the existence of the Fréchet p -mean is ensured if the MpD is finite at one point, thus at all points thanks to the triangle inequality. The uniqueness is a much more difficult problem. For smooth differential geometric spaces like Riemannian spaces, and restricting to the classical mean with $p = 2$, it has been argued in [2, p.235] that *the existence of a unique center of mass in the large for manifolds with non-positive curvature was proven and used by Élie Cartan back in the 1920's*. In order to find the fixed point of a group of isometries, Cartan indeed showed in [4]² that the sum of the square distance to a finite number of points has a unique minimum in simply connected Riemannian manifolds with non-positive curvature (now called Hadamard spaces). This result was extended in [5] to closed subgroups of isometries³: *”Let us apply to the point at origin O the displacements defined by the transformations of γ . The group γ being closed, we obtain a closed manifold V (that can be reduced to a point). But in a Riemann space without singular point at finite distance, simply connected, with negative or zero curvature, given points in finite number, we can find a fixed point invariant by all the displacements that exchanges these points: this is the point for which the sum of square distances to the given point is minimal ([4, p.267]). This property is still true if, instead of a finite number of points, we have an infinite number forming a closed manifold: we arrive at the conclusion that the group γ , which leave evidently the manifold V invariant, also leaves invariant a fixed point of the space. Thus, this point belongs to the group or (isometric) rotations around this point. But this group is homologous to g in the continuous adjoint group, which demonstrate the theorem.”* It is obvious in this text that Cartan is only using the uniqueness of the sum of square distances as a tool in the specific case of negative curvature Riemannian manifolds and not as a general definition of the mean on manifolds as is intended in probability or statistics.

In 1973, for similar purposes, Grove and Karcher extended Cartan's idea to positive curvature manifolds. However, they restricted their definition to distribution with sufficiently small support so that the mean exists and is unique [18]. The notion was coined as *the Riemannian center of mass*. In this publication and in successive ones, Karcher and colleagues used Jacobi field estimates to determine the largest convex geodesic ball that support this definition. The Riemannian barycenter is commonly refereed to as being introduced in [22]. However, the most complete description of the related properties is certainly found in [3], where a notion of barycenter in affine connection manifolds is also worked out. A good historical note on the history of the Riemannian barycenter is given in [1] and by Karcher himself in [23].

Exponential barycenters In all the above works, the Riemannian center of mass is by definition unique. Considering a set-valued barycenter on an affine connection manifold was the contribution of Emery and Mokobodzki [11]. In a Riemannian manifold, at the points $x \in \mathcal{M}$ where the cut locus has null measure for the measure μ , the critical points of the mean square

¹Page 259: ”It is not certain that such an element exists nor that it is unique.”

²Note III on normal spaces with negative or null Riemannian curvature, p.267.

³Appliquons au point origine O les différents déplacements définis par les transformations de γ . Le groupe γ étant clos, nous obtenons ainsi une variété fermée V (qui peut se réduire à un point). Or, dans un espace de Riemann sans point singulier à distance finie, simplement connexe, a courbure negative ou nulle, on peut trouver, étant donnés des points en nombre fini, un point fixe invariant par tous les déplacements qui échangent entre eux les points donnés : c'est le point pour lequel la somme des carrés des distances au point donné est minima ([4, p.267]). Cette propriété est encore vraie si, au lieu d'un nombre fini de points, on en a une infinité formant une variété fermée : nous arrivons donc à la conclusion que le groupe γ qui laisse évidemment invariante la variété V , laisse invariant un point fixe de l'espace, il fait donc partie du groupe des rotations (isométriques) autour de ce point. Mais ce groupe est homologue à g dans le groupe adjoint continu, ce qui démontre le théoreme.

distance are characterized by the barycentric equation:

$$\mathfrak{M}_1(x) = \int_{\mathcal{M}} \log_x(y) d\mu(y) = 0, \quad (2)$$

where $\log_x(z)$ is the initial tangent vector of the minimizing geodesic joining x to z (the Riemannian log). This barycentric equation was taken as the definition of *exponential barycenters* of the probability measure [11] in more general affine connection spaces. Notice that the notion remains purely affine, provided that the distribution has a support on a convex neighborhood in which the logarithm can be defined uniquely. The non-uniqueness of the expectation of a random variable considerably extended the usability of this definition, in particular in positive curvature spaces. In the Riemannian case, exponential barycenters contain in particular the smooth local (and the global) minimizers of the means square distance (MSD) $\sigma^2(x) = \int_M \text{dist}^2(x, y) \mu(dy)$ (except those at which the MSD is not differentiable). Exponential barycenters were later used in [35, 7] and in [40, 31, 43] for bi-invariant means on Lie groups endowed with the canonical symmetric Cartan-Schouten connection.

Uniqueness of the Fréchet mean Conditions for the uniqueness of the minimum of the sum of square distance have been studied by Karcher [22, 3] and further optimized in [25, 27, 28].

Theorem 1 (Karcher & Kendall Concentration (KKC) Conditions) *Let \mathcal{M} be a geodesically complete Riemannian manifold with injection radius $\text{inj}(x)$ at x . Let μ is a probability distribution on \mathcal{M} whose support is contained a closed regular geodesic ball $\bar{B}(x, r)$ of radius $r < \frac{1}{2}\text{inj}(x)$. We assume moreover that the upper bound $\kappa = \sup_{x \in B(x, r)}(\kappa(x))$ of the sectional curvatures in that ball satisfies $\kappa < \pi^2/(4r)^2$. This second condition is always true on spaces of negative curvature and specifies a maximal radius $r^* = \frac{\pi}{2\sqrt{\kappa}}$ when there is positive sectional curvature. These concentration assumptions ensure that the MSD has a unique global minimum that belongs to the ball $\bar{B}(x, r)$.*

The result has been extended to Fréchet p -means defined as the minimizers of the mean p -distance in [1, 51, 52].

In order to differentiate the different notions of means in Riemannian manifolds, it has been agreed in geometric statistics to name Fréchet mean the set of global minimizers of the MSD, Karcher mean the set of local minimizers, and exponential barycenters the set of critical points satisfying the implicit equation $\mathfrak{M}_1(x) = 0$. It is clear that all these definition boils down for the classical 2-mean to the same unique point within the ball $B(x, r)$ of the KKC conditions. Notice that we may still have local minima and critical points located outside this ball.

3 Statistics beyond the mean value: generalizing PCA

The mean value is just a zero-dimensional summary statistics of the data. If we want to retain more information, we need to add more degrees of freedom, and Principal Component Analysis (PCA) is the ubiquitous tool for that. In a Euclidean space, the principal k -dimensional affine subspace of the 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.

Tangent PCA Generalizing PCA to manifolds and to potentially more general stratified spaces is currently a very active research topic. The first step is the generalization of affine subspaces in manifolds. For the zero-dimensional subspace, the Fréchet mean is the natural intrinsic

generalization of the mean around which PCA is performed. 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 [37]. 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).

Principal Geodesic Analysis (PGA) Instead of an analysis of the covariance matrix, [13] 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. In fact, the tangent vectors also need to be restricted to the interior of the tangential cut locus within this linear subspace if we want to generate a submanifold of the original manifold [39]. The idea of minimizing the unexplained variance (i.e. the norm of the residuals) is really interesting and corresponds exactly to what we want to do in manifold dimension reduction. However, the non-linear least-squares procedure to optimize the geodesic subspace is computationally expensive, so that [13] approximated in practice PGA with tPCA. This is really unfortunate because this led many people to confuse PGA and tPCA. A real implementation of the original PGA procedure was only recently provided by [48]. 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.

Geodesic PCA In the PGA procedure, the mean always belongs to geodesic subspaces even when it is outside of the distribution support. To alleviate this problem, Huckemann et al. [20, 19] proposed to start at the first order component directly with the geodesic that best fits the data, which is not necessarily going through the mean. The second principal geodesic is chosen orthogonally at a point of 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 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 [47].

Principal Nested Spheres All the previous extensions of PCA are intrinsically forward methods that build successively larger approximation spaces for the data. A notable exception to this principle is the concept of Principal Nested Spheres (PNS), proposed by [21] 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 have recently generalized this approach to manifolds with the help of a “nested sequence of relations” [6]. However, such a sequence was only known so far for spheres or Euclidean spaces.

3.1 Barycentric subspaces in manifolds

The geodesic subspaces used for tPCA, PGA and GPCA are described by one point and k tangent vectors at that point. This give a special role to this reference point which is not found in the higher dimensional descriptors. Moreover, these spaces are totally geodesic at the reference point but generally nowhere else. This asymetry may not be optimal for multimodal distributions that do not have a single 'pole'.

Fréchet, Karcher and Exponential barycentric subspaces In order to have a fully symmetric and 'multi-pole' description of subspaces, we have proposed in [39] a new and more general type of subspaces in manifolds: *barycentric subspaces* are implicitly defined as the locus of weighted means of $k + 1$ reference points with positive or negative weights summing up to one. This time the descriptors are fully symmetric (they are all standard points) Depending on the definition of the mean, we obtain the Fréchet, Karcher or exponential barycentric subspaces (FBS/KBS/EBS). The Fréchet (resp. Karcher) barycentric subspace of the points $(x_0, \dots, x_k) \in \mathcal{M}^{k+1}$ is the locus of weighted Fréchet (resp. Karcher) means of these points, i.e. the set of global (resp. local) minimizers of the weighted MSD: $\sigma^2(x, \lambda) = \frac{1}{2} \sum_{i=0}^k \lambda_i \text{dist}^2(x, x_i)$:

$$\text{FBS}(x_0, \dots, x_k) = \left\{ \arg \min_{x \in \mathcal{M}} \sigma^2(x, \lambda), \lambda \in R^{k+1}, \mathbf{1}^\top \lambda = 1 \right\}.$$

The EBS is the locus of weighted exponential barycenters of the reference points (critical points of the weighted MSD) defined outside their cut-locus as follows.

Definition 2 (Exponential Barycentric Subspace (EBS) and Affine Span) *A set of $k+1$ points $\{x_0, \dots, x_k\} \in \mathcal{M}^{k+1}$ is affinely independent if no point is in the cut-locus of another and if all the $k+1$ 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.*

The EBS of $k+1$ affinely independent points (x_0, \dots, x_k) is the locus of weighted exponential barycenters of the reference points:

$$\text{EBS}(x_0, \dots, x_k) = \{x \in \mathcal{M} \setminus \text{Cut}(x_0, \dots, x_k) \mid \exists \lambda \in R^{k+1}, \mathbf{1}^\top \lambda = 1 : \sum_i \lambda_i \log_x(x_i) = 0\}.$$

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.

Thus, outside the cut locus of the reference points, we clearly see the inclusion $\text{FBS} \subset \text{KBS} \subset \text{EBS}$. The completeness of the affine span allows ensuring that a closest point exists on the subspace, which is fundamental in practice for optimizing the subspaces by minimizing the distance of the data to their projection. This definition works on metric spaces more general than Riemannian manifolds. In stratified metric spaces, the barycentric subspace spanned by points belonging to different strata naturally maps over several strata [50].

Barycentric subspaces can be characterized very similarly to the Euclidean case by the singular value decomposition of a certain matrix or by the diagonalization of the covariance and the Gram matrices. Let $Z(x) = [\log_x(x_0), \dots, \log_x(x_k)]$ be the matrix field of the log of the reference points x_i in a local coordinate system. This is a smooth field outside the cut locus of the reference points. The EBS is the zero level-set of the smallest singular value of $Z(x)$. The associated right singular vector gives the weights λ that satisfy the barycentric equation $\sum_i \lambda_i \log_x(x_i) = 0$. Denoting $G(x)$ the matrix expression of the Riemannian metric, we can also define the smooth $(k+1) \times (k+1)$ Gram matrix field $\Omega(x) = Z(x)^\top G(x) Z(x)$ with components $\Omega_{ij}(x) = \langle \overrightarrow{xx_i} \mid \overrightarrow{xx_j} \rangle_x$ and the (scaled) $n \times n$ covariance matrix field of the reference points $\Sigma(x) = \sum_{i=0}^k \overrightarrow{xx_i} \overrightarrow{xx_i}^\top = Z(x) Z(x)^\top$. With these notations, $\text{EBS}(x_0, \dots, x_k)$ 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 $\Sigma(x)$.

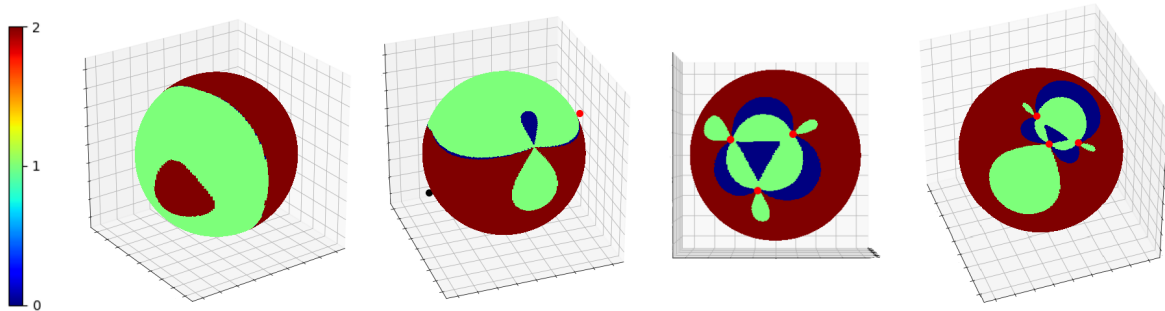


Figure 1: Signature of the weighted Hessian matrix for different configurations of 3 reference points (in black, antipodal point in red) on the 2-sphere: the locus of local minima (KBS) in brown does not cover the whole sphere and can even be disconnected (first example). Reproduced from [39]

Example in constant curvature spaces It is interesting to look at the barycentric subspaces in the most simple non-linear spaces: constant curvature spaces. The sphere can be represented by the classical embedding of the unit sphere $\mathcal{S}_n \subset \mathbb{R}^{n+1}$ and the hyperbolic space as the unit pseudo-sphere of the Minkowski space $\mathbb{R}^{1,n}$. With this model, the affine span of $k + 1$ reference points is particularly simple: it is the great subsphere (resp great sub-pseudosphere) obtained by the intersection of the (pseudo) sphere with the hyperplane generated by the reference points in the embedding space. Notice that the EBS in the sphere case is the great subsphere (the affine span) minus the antipodal points (the cut locus) of the reference points. However, even if the affine span is simple, understanding the Fréchet / Karcher barycentric subspaces is much more involved. In order to find the local minima among the critical points, we have to compute the Hessian matrix of the weighted MSD, and to look at the sign of its eigenvalues: at the critical points with a non-degenerate Hessian (regular points), local minima are characterized by a positive definite Hessian. In fact, the zero-eigenvalues of the Hessian subdivide the EBS into a cell complex according to the index (the number of positive eigenvalues) of the Hessian. This index is illustrated on Fig. (1) for a few configuration of 3 affinely independent reference points on the 2-sphere: we clearly see that 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 shows that the affine span is a more interesting definition than KBS / FBS.

Properties of barycentric subspaces For $k + 1$ affinely independent reference points, the affine span is a smooth submanifold of dimension k in a neighborhood of each of the reference points. Further away from the reference points, one can show that the EBS is a stratified spaces of maximal dimension k at the regular points [39]. At the degenerate points where the Hessian has null eigenvalues, pathological behavior with an increasing dimension may perhaps appear although this has not been observed in constant curvature spaces.

When the reference points belong to a sufficiently small regular geodesic ball (typically satisfying the KKC conditions of Theorem 1), then the FBS with positive weights, called the barycentric simplex, is the graph of a k dimensional function. The barycentric k -simplex contains all the reference points and the geodesics segments between the reference points. The $(k - l)$ -faces of a 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. Barycentric simplexes were investigated in [50] as extensions of principal subspaces in the negatively curved metric spaces of trees under the name Locus of Fréchet mean (LFM).

Making the $k + 1$ reference points become closer together, we may imagine that at the limit

they coalesce at one point x along k directions $(v_1, \dots, v_k) \in T_x \mathcal{M}^k$. In that case, one can show that the EBS converges to the geodesic subspace generated by geodesics starting at x with tangent vectors that are linear combination of the vectors (v_1, \dots, v_k) (up to the cut locus of x). More generally, the reference points may converge to a non local jet⁴, which may give a new way to define multi-dimensional splines in manifolds.

3.2 From PCA to barycentric subspace analysis

The nestedness of approximation spaces is one of the most important characteristics for generalizing PCA to more general spaces [6]. Barycentric subspaces can easily be nested by adding or removing one or several points at a time, which corresponds to put the barycentric weight of this (or these) point(s) to zero. This gives a family of embedded submanifolds called a flag because this generalizes flags of vector spaces.

Forward and backward approaches With a forward analysis, we compute iteratively the flag of affine spans by adding one point at a time keeping the previous ones fixed. Thus, we begin by computing the optimal barycentric subspace $\text{Aff}(x_0) = \{x_0\}$, which is the set of exponential barycenters. This may be a Fréchet mean or more generally a critical value of the unexplained variance. Adding a second point amounts to computing the geodesic passing through the mean that best approximates the data. Adding a third point now generally differs from PGA. In practice, the forward analysis should be stopped at a fixed number or when the variance of the residuals reaches a threshold (typically 5% of the original variance). We call this method the forward barycentric subspace (FBS) decomposition of the manifold data. Due to the greedy nature of this forward method, the affine span of dimension k defined by the first $k + 1$ points is not in general the optimal one minimizing the unexplained variance.

The backward analysis consists in iteratively removing one dimension. Starting with a set of points x_0, \dots, x_n which generates the full manifold, we could start to chose which one to remove. However, as the affine span of $n + 1$ linearly independent points generate the full manifold, the optimization really begins with n points. Once they are fixed, the optimization boils down to test which point should be removed. In practice, we may rather optimize $k + 1$ points to find the optimal k -dimensional affine span, and then reorder the points using a backward sweep to find inductively the one that least increases the unexplained variance. We call this method the k -dimensional pure barycentric subspace with backward ordering (k-PBS). With this method, the k -dimensional affine span is optimizing the unexplained variance, but there is no reason why any of the lower dimensional ones should.

Barycentric Subspace Analysis: optimizing a criterion on the whole flag of subspaces

In order to obtain optimal subspaces which are embedded 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. This idea gives the Accumulated Unexplained Variances (AUV) criterion [39], which sums the unexplained variance (the sum of squared norm of the residuals) by all the subspaces of the flag. AUV is analogous to the Area-Under-the-Curve (AUC) in Receiver Operating Characteristic (ROC) curves. In practice, one can stop at a maximal dimension k like for the forward analysis in order to limit the computational complexity. This analysis limited to a flag defined by $k + 1$ points is denoted

⁴ 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 . Non-local jets, or multijets, generalize subspaces of the tangent spaces to higher differential orders with multiple base points.

k -BSA. Let us denote by $\sigma_{out}^2(\text{Aff}(x_0, \dots, x_k))$ the sum of the squared distance of the data points to the affine subspace $\text{Aff}(x_0, \dots, x_k)$ (the variance unexplained by this subspace). The AUV of the flag $\text{Aff}(x_0) \subset \text{Aff}(x_0, x_1) \subset \dots \subset \text{Aff}(x_0, \dots, x_k)$ is

$$AUV(x_0, x_1, \dots, x_k) = \sigma_{out}^2(\text{Aff}(x_0)) + \sigma_{out}^2(\text{Aff}(x_0, x_1)) + \dots + \sigma_{out}^2(\text{Aff}(x_0, \dots, x_k)). \quad (3)$$

In a Euclidean space, minimizing the unexplained variance with respect to a k -dimensional affine subspace leads to select the hyperplane going through the mean of the data and spanned by the eigenvectors (v_1, \dots, v_k) of the covariance matrix associated to the smallest k eigenvalues. However, this subspace is independent of the ordering of the k selected eigenvectors, and nothing in this criterion helps us to select an optimal smaller dimensional subspace. Thus, the PCA ordering of the subspaces that produces a hierarchy of embedded subspaces of larger dimension is not specified by the unexplained variance criterion (nor by the explained variance) alone. If we now consider the AUV criterion on a flag of subspaces (say up to order k), then one can show that the unique solution is the flag generated with the increasing ordering of the eigenvalues. This shows that PCA is actually optimizing the AUV criterion on the space of affine flags. The generalization to the optimization of the AUV criterion on the space of flags of affine spans in Riemannian manifolds is called Barycentric Subspaces Analysis (BSA).

Forward, backward and optimal flag estimation In summary, we may consider three main algorithms to estimate a flag of barycentric subspaces of maximal dimension k :

- The Forward Barycentric Subspace decomposition (k -FBS) iteratively adds the point that minimizes the unexplained variance up to $k + 1$ points. With this method, only the first (usually 0 or 1 dimensional) subspace optimized is optimal and all the larger ones are suboptimal for the unexplained variance.
- The optimal Pure Barycentric Subspace with backward reordering (k -PBS) estimates the $k + 1$ points that minimize the unexplained variance for a subspace of dimension k , and then reorders the points accordingly for lower dimensions. Here only the largest subspace is optimal, and all the smaller ones are suboptimal.
- The Barycentric Subspace Analysis of order k (k -BSA) looks for the flag of affine spans defined by $k + 1$ ordered points that optimized the AUV. Here none of the subspaces are optimal for the unexplained variance of the corresponding dimension, but the whole flag is optimal for the AUV criterion.

3.3 Sample-limited L_p barycentric subspace inference

Sample-limited inference In order to compute the optimal subspaces or flags of subspaces, we need to set-up an optimization procedure. This can be realized by standard gradient descent techniques for optimization on manifolds. However, to avoid gradient computations but also to avoid finding optima that are far away from any data point, it has been proposed to limit the inference of the Fréchet mean to the data-points only. For instance, in neuroimaging studies, the individual image minimizing the sum of square deformation distance to other subject images was found to be a good alternative to the mean template (a Fréchet mean in deformation and intensity space) because it conserves the original characteristics of a real subject image [30]. Beyond the Fréchet mean, Feragen et al. proposed to define the first principal component mode as the unexplained variance minimizing geodesic going through two of the data points [12]. The method named *set statistics* was aiming to accelerate the computation of statistics on tree spaces. [53] further explored this idea under the name of *sample-limited geodesics* in the context of PCA in phylogenetic tree space. In both cases, defining principal modes of order larger than two was seen as an unsolved challenging research topic.

With barycentric subspaces, the idea of sample-limited statistics naturally extends to any dimension by restricting the search to the (flag of) affine spans that are parametrized by points sampled from the data. The implementation boils down to an enumeration problem. With this technique, the reference points are never interpolated as they are by definition sampled from the data. This is an important advantage for interpreting the modes of variation since we may go back to other information about the samples like the medical history and disease type. The search can be done exhaustively for a small number of reference points. The main drawback is the combinatorial explosion of the computational complexity with the dimension for the optimal order- k flag of affine spans, which is involving $O(N^{k+1})$ operations, where N is the number of data points. In [38] we perform an exhaustive search, but approximate optima can be sought using a limited number of randomly sampled points [12].

Adding robustness with L_p norms Since barycentric subspaces minimize the weighted MSD, one could think of taking a power p of the metric to define the mean p -distance (MpD) $\sigma^p(x) = \frac{1}{p} \sum_{i=0}^k \text{dist}^p(x, x_i)$. We recall that the global minimizers of this criterion defines the Fréchet median for $p = 1$, the Fréchet mean for $p = 2$ and the barycenter of the support of the distribution for $p = \infty$. This suggests to further generalize barycentric subspaces by taking the locus of the minima of the weighted unexplained MpD $\sigma^p(x, \lambda) = \frac{1}{p} \sum_{i=0}^k \lambda_i \text{dist}^p(x, x_i)$. However, it turns out that the critical points of the weighted MpD are necessarily included in the EBS since the gradient of the MpD is (except at the reference points and their cut loci):

$$\nabla_x \sigma^p(x, \lambda) = \nabla_x \frac{1}{p} \sum_{i=0}^k \lambda_i (\text{dist}^2(x, x_i))^{p/2} = - \sum_{i=0}^k \lambda_i \text{dist}^{p-2}(x, x_i) \log_x(x_i).$$

Thus, we see that the critical points of the MpD 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 also elements of the EBS and changing the power of the metric just amounts to a reparameterization of the barycentric weights. This stability of the EBS / affine span with respect to the power of the metric p shows that the affine span is really a central notion.

While changing the power does not change the subspace definition, it has a drastic impact on its estimation: minimizing the sum of L_p distance to the subspace for non-vanishing residuals obviously changes the relative influence of points [38]. It is well known that medians are more robust than least-squares estimators: the intuitive idea is to minimize the power of residuals with $1 \leq p \leq 2$ to minimize the influence of outliers. For $0 < p < 1$, the influence of the closest points becomes predominant, at the cost of non-convexity. In general, this is a problem for optimization. However, there is no gradient estimation in the sample-limited setting as we have to rely on an exhaustive search for the global minimum or on a stochastic approximation by testing only a subset of reference configurations. At the limit of $p = 0$, all the barycentric subspaces containing $k + 1$ points (i.e. all the sample-limited barycentric subspaces of dimension k that we consider) have the same L_0 sum of residuals, which is a bit less interesting.

For a Euclidean space, minimizing the sum of L_p norm of residuals under a rank k constraint is essentially the idea of the robust R1-PCA [8]. However, an optimal rank k subspace is not in general a subspace of the optimal subspace of larger ranks: we lose the nestedness property. An alternative PCA-L1 approach, which maximizes the L_1 dispersion within the subspace, was proposed in [26]. On manifolds, this would lead to a generalization of tangent-PCA maximizing the explained p -variance. In contrast, we proposed in [38] to minimize the Accumulated Unexplained p -Variance (L_p AUV) over all the subspaces of the flag under consideration. Since the subspace definition is not affected by the power p , we can compare the subspaces' parameters (the reference points) for different powers. It also allows to simplify the algorithms: as the (positive) power of a (positive) distance is monotonic, the closest point to an affine span for the 2-distance remains the closest point for the p -distance. Thus, the forward barycentric subspace analysis (k -FBS), the pure subspace with backward reordering analysis (k -PBS) and the

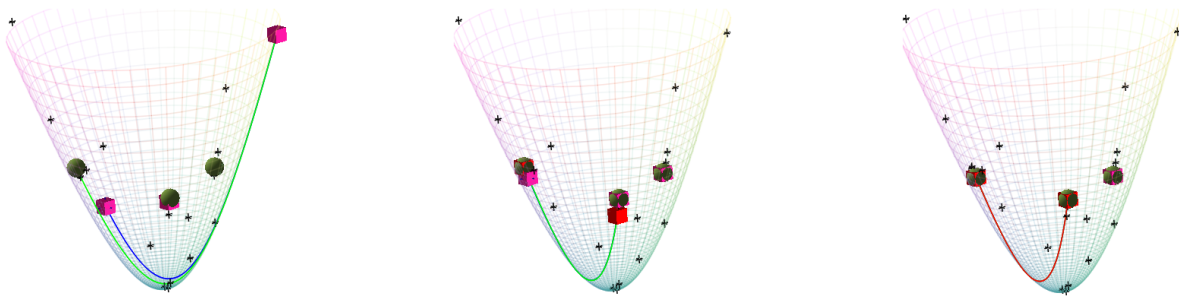


Figure 2: Analysis of 3 clusters on a 5D hyperbolic space, projected to the expected 2-pseudo-sphere, with $p = 2$ (**left**), $p = 1$ (**middle**) and $p = 0.5$ (**right**). For each method (FBS in blue, 1-PBS in green, 1-BSA in red), the 1d mode is figured as a geodesic joining the two reference point (unseen red, blue or green geodesics are actually hidden by another geodesic). The three reference points of 2-PBS are represented with dark green solid circles, and the ones of 2-BSA with deep pink solid boxes. Reproduced, with permission from [38].

barycentric subspace analysis (k -BSA) can be seamlessly generalized to their robust L_p version in the sample-limited setting.

4 Example applications of Barycentric subspace analysis

4.1 Example on synthetic data in a constant curvature space

The L_p variant of the forward (FBS), backward (PBS) and BSA algorithms were evaluated on synthetically generated data on spheres and hyperbolic spaces in [38]. The projection of a point of a sphere on a subsphere is almost always unique and corresponds to the renormalization of the projection on the Euclidean subspace containing the subsphere. The same property can be established for hyperbolic spaces, which can be viewed as pseudo-spheres embedded in a Minkowski space. Affine spans are great pseudo-spheres (hyperboloids) generated by the intersection of the plane containing the reference points with the pseudo-sphere, and the closest point on the affine span is the renormalization of the unique Minkowski projection on that plane [39]. In both cases, implementing the Riemannian norm of the residuals is very easy and the difficulty of sample-limited barycentric subspace algorithms analysis resides in the computational complexity of the exhaustive enumeration of tuples of points.

We illustrate in Fig.2 the results of the L_p barycentric subspace algorithms on a set of 30 points in the 5D hyperbolic space generated at follows: we draw 5 random points (tangent Gaussian with variance 0.015) around each vertex of an equilateral triangle of length 1.57 centered at the bottom of the 5D hyperboloid embedded in the (1,5)-Minkowski space. As outliers, we add 15 points drawn according to a tangent Gaussian of variance 1.0 truncated at a maximum distance of 1.5 around the bottom of the 5D hyperboloid. This simulates three clusters living on a lower dimensional 2-pseudo-sphere with 50% of outliers (Fig.2). With the L_2 hyperbolic distance, the 1-FBS and 1-BSA methods select outliers for their two reference points. 1-PBS manages to get one point in a cluster. For the two dimensional approximation again with the L_2 distance, the 2-FBS and the 2-PBS select only one reference points within the clusters while 2-BSA correctly finds the clusters (Fig.2 left, dark green points). With the L_1 distance, FBS, PBS and BSA select 3 very close points within the three clusters (Fig.2 center). Lowering the power to $p = 0.5$ leads to selecting exactly the same points optimally centered within the 3 clusters for all the methods (Fig.2 right). Thus, it seems that we could achieve some kind of *Principal Cluster Analysis* with the sample-limited L_p barycentric subspace analysis.

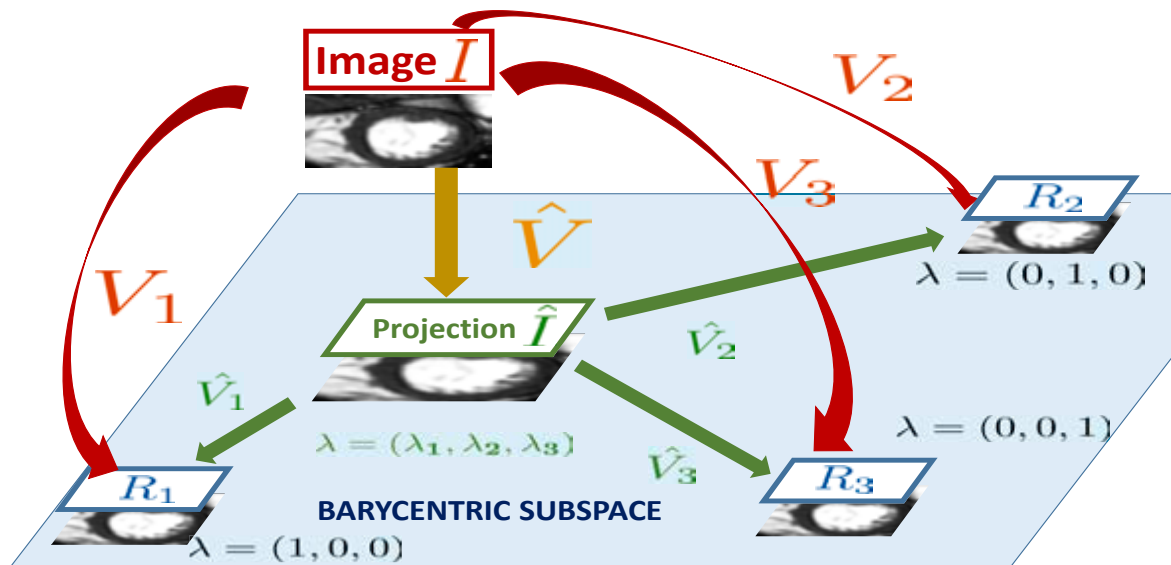


Figure 3: Barycentric subspace of dimension 2 built from 3 reference images (R_1, R_2, R_3) . \hat{I} is the projection of the image I within the barycentric subspace such that $\|\hat{v}\|^2$ is minimal under the conditions $\sum_j \lambda_j \hat{V}_j = 0$ and $\hat{V} + \hat{V}_j = V_j$. Reproduced from [45].

4.2 A symmetric group-wise analysis of cardiac motion in 4D image sequences

Understanding and analyzing the cardiac motion pattern in a patient is an important task in many clinical applications. The cardiac motion is usually studied by finding correspondences between each of the frames of the sequence and the first frame corresponding to the end-diastole (ED) image. This image registration process yields a dense displacement field that tracks the motion of the myocardium over the image sequence. Taking the ED image as a reference is natural as it is the starting point of the contraction of the heart which is the most important phase in evaluating the efficiency of the cardiac function. However, this specific choice can lead to important biases in quantifying the motion, especially at end-systole (ES) where the deformations to be evaluated are large. In [45], we proposed to build a multi-reference registration to a barycentric subspaces of the space of images representing cardiac motion instead of taking a unique reference image to evaluate the motion.

In the context of diffeomorphic medical image registration, 3D images are the “points” of our manifold while “geodesics” are the optimal deformations found by image registration to map one image to the other. In the Large Diffeomorphic Metric Mapping (LDDMM) as well as in the Stationary Velocity Field (SVF) registration frameworks, diffeomorphic deformations are obtained by the flow of velocity fields, and the tangent vector (i.e. a vector field over the image) at the initial point of the deformation trajectory registering image I to image J may be interpreted as $\log_I(J)$. We refer the reader to [43, 31] for a more in-depth description of the SVF framework and its application in medical image registration.

The barycentric subspace of dimension k spanned by $k+1$ reference images R_1 to R_{k+1} is then defined as the set of images \hat{I} for which there exists weights λ_i such that $\sum_{j=1}^{k+1} \lambda_j \log_j(R_j) = 0$. Thus, projecting image I on this subspace amounts to find the smallest SVF \hat{V} deforming image I to image \hat{I} such that the SVFs \hat{V}_1, \hat{V}_2 and \hat{V}_3 encoding the deformation from this projected image to the three references R_1, R_2 and R_3 are linearly dependent (Figure 3). The weights λ_j are the barycentric coordinates of image I in the “basis” (R_1, R_2, R_3) . This process can be repeated for each image I_1 to I_N of the temporal sequence of one subject. This allows us to

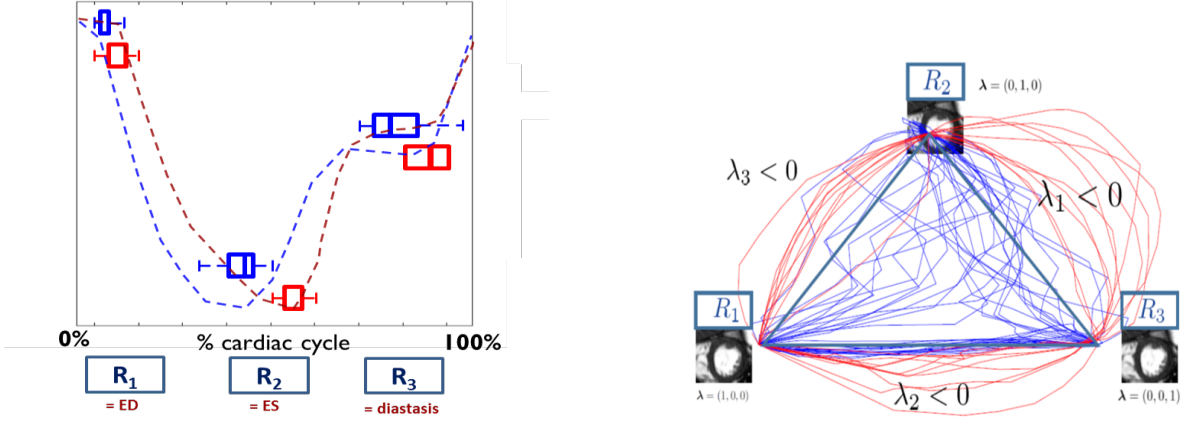


Figure 4: Cardiac motion signature of 10 sequences of control subjects (in blue) and 16 sequences of Tetralogy of Fallot subjects (in red). Left: the time-index of the three optimal references superimposed on the average cardiac volume curve for each population. Right: the curve of barycentric coordinates of the images along the sequence of each subject, projected on the plane $\sum_i \lambda_i = 1$. Modified from [45, 44].

compute the unexplained variance $\sigma^2(R_1, R_2, R_3) = \sum_{i=1}^N \|\hat{V}(I_i)\|^2$ and to choose the optimal basis by repeating the experiment for all possible triplets of reference images within our sequence.

This methodology was used in [45] to compare the cardiac motion signature of two different populations. The first group consists of 10 healthy subjects from the STACOM 2011 cardiac motion tracking challenge dataset [49], and the second group is made of 10 Tetralogy of Fallot (ToF) patients [33]. Short axis cine MRI sequences were acquired with $T = 15$ to 30 frames. For each subject, we projected each of the frames of the cardiac sequence to a barycentric subspace of dimension 2 built by 3 reference images belonging to that sequence. A significant differences in the time-index of the optimal references frames can be seen between the two populations (Fig. 4, left). In particular, the second reference - corresponding to the end-systole - is significantly later for the ToF patients showing that this population has on average longer systolic contraction. The barycentric coefficients (Fig. 4, right) also show significant differences between the groups, especially in the region $\lambda_1 < 0$. This demonstrates that this signature of the motion is encoding relevant features. Last but not least, the reconstruction based on the 3 reference images and the 30 barycentric coefficients along the sequence (a compression rate of 1/10) turned out to achieve a reconstruction error in intensity which is 40% less than the one of a classical tangent PCA with two deformation modes at the mean image of the sequence (compression rate 1/4 only). This demonstrates that the multi-reference approach based on barycentric subspaces can outperform the classical statistical shape analysis methods on real medical imaging problems.

5 Conclusion and perspectives

We have drafted in this chapter a summary of some of the recent advances on manifold dimension reduction and generalization of principal component analysis to Riemannian manifolds. The first observation is that generalizing affine subspaces to manifolds with a minimization procedure (i.e. Fréchet or Karcher barycentric subspaces) leads to small disconnected patches that do not cover complete lower dimensional subspheres (resp. sub-pseudospheres) in constant curvature spaces. Considering the completion (the affine span) of all critical points (the exponential barycentric subspace) is needed to cover the full sub-(pseudo)-sphere. The fact that changing the mean square distance for the mean p -distance does not change the affine span is an unexpected stability result which suggests that the notion is quite central. The second important point is that geodesic subspaces that were previously proposed for PGA or GPCA are actually particular cases of

barycentric subspaces that can be obtained by taking reference points highly concentrated with respect to the distribution of our data points. The framework is thus more general. Last but not least, following [6], any interesting generalization of PCA should rely on sequences of properly nested spaces. Generalizing linear flags in Euclidean space, an ordering of the reference points naturally defines a flag of nested affine spans in a manifold. Now instead of defining the subspaces of the flag iteratively in a forward or backward manner, which is sub-optimal for each subspace considered independently, it turns out that PCA can be rephrase as an optimization of the AUV criterion (the sum of all unexplained variances by all the subspaces of the hierarchy) on the space of flags. Such a method coined Barycentric Subspace Analysis can be naturally extended to the L_p norm of residuals to account for outliers or different type of noises. BSA can also be performed by restricting reference points defining the subspaces to be a subset of the data points, thus considerably extending the toolbox of sample-limited statistics to subspaces of dimension larger than one, and an example application on 3D image sequences of the heart showed that many insights ca be brought by this new methodology.

Acknowledgments: This work has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant G-Statistics No 786854).

References

- [1] B. Afsari. Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(02):655–673, Feb. 2011.
- [2] M. Berger. *A Panoramic View of Riemannian Geometry*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.
- [3] P. Buser and H. Karcher. *Gromov’s almost flat manifolds*. Number 81 in Astérisque. Société mathématique de France, 1981.
- [4] E. Cartan. *Leçons sur la géométrie des espaces de Riemann*. Gauthier-Villars, Paris, France, 1928.
- [5] E. Cartan. Groupes simples clos et ouverts et géométrie riemannienne. *Journal de Mathématiques Pures et Appliquées*, 9e série(tome 8):1–34, 1929.
- [6] J. Damon and J. S. Marron. Backwards Principal Component Analysis and Principal Nested Relations. *Journal of Mathematical Imaging and Vision*, 50(1-2):107–114, October 2013.
- [7] R. W. R. Darling. Geometrically intrinsic nonlinear recursive filters II: Foundations, 1998. arXiv:math/9809029.
- [8] C. Ding, D. Zhou, X. He, and H. Zha. R1-pca: Rotational invariant l1-norm principal component analysis for robust subspace factorization. In *Proceedings of the 23rd International Conference on Machine Learning, ICML ’06*, pages 281–288, New York, NY, USA, 2006. ACM.
- [9] I. Dryden and K. Mardia. Theoretical and distributional aspects of shape analysis. In *Probability Measures on Groups, X (Oberwolfach, 1990)*, pages 95–116, New York, 1991. Plenum.
- [10] A. Edelman, T. Arias, and S. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal of Matrix Analysis and Applications*, 20(2):303–353, 1998.

- [11] M. Emery and G. Mokobodzki. Sur le barycentre d'une probabilité dans une variété. In *Séminaire de Probabilités XXV*, volume 1485, pages 220–233. Springer, Berlin, Heidelberg, 1991.
- [12] A. Feragen, M. Owen, J. Petersen, M. M. W. Wille, L. H. Thomsen, A. Dirksen, and M. de Bruijne. Tree-space statistics and approximations for large-scale analysis of anatomical trees. In J. C. Gee, S. Joshi, K. M. Pohl, W. M. Wells, and L. Zöllei, editors, *Information Processing in Medical Imaging*, pages 74–85, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- [13] P. Fletcher, C. Lu, S. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, Aug. 2004.
- [14] M. Fréchet. Valeurs moyennes attachées a un triangle aléatoire. *La revue scientifique*, Fascicule 10:475–482, Dec. 1943.
- [15] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, 10:215–310, 1948.
- [16] C. Gramkow. On averaging rotations. *International Journal of Computer Vision*, 42(1-2):7–16, April/May 2001.
- [17] U. Grenander, M. Miller, and A. Srivastava. Hilbert-schmidt lower bounds for estimators on matrix Lie groups for ATR. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 20(8):790–802, 1998.
- [18] K. Grove and H. Karcher. How to conjugate C^1 -close group actions. *Mathematische Zeitschrift*, 132(1):11–20, Mar. 1973.
- [19] S. Huckemann, T. Hotz, and A. Munk. Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions. *Statistica Sinica*, 20:1–100, 2010.
- [20] S. Huckemann and H. Ziezold. Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces. *Advances in Applied Probability*, 38(2):299–319, June 2006.
- [21] S. Jung, I. L. Dryden, and J. S. Marron. Analysis of principal nested spheres. *Biometrika*, 99(3):551–568, September 2012.
- [22] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, Sept. 1977.
- [23] H. Karcher. Riemannian Center of Mass and so called Karcher mean. *arXiv:1407.2087 [math]*, July 2014.
- [24] D. Kendall. A survey of the statistical theory of shape (with discussion). *Statistical Science*, 4:87–120, 1989.
- [25] W. Kendall. Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. *Proc. London Math. Soc.*, 61(2):371–406, 1990.
- [26] N. Kwak. Principal component analysis based on l1-norm maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9):1672–1680, Sept 2008.
- [27] H. Le. Locating Fréchet means with application to shape spaces. *Advances in Applied Probabilities*, 33:324–338, 2001.

- [28] H. Le. Estimation of Riemannian barycenters. *LMS J. Comput. Math.*, 7:193–200, 2004.
- [29] H. Le and D. Kendall. The Riemannian structure of Euclidean shape space: a novel environment for statistics. *Annals of Statistics*, 21:1225–1271, 1993.
- [30] N. Leporé, C. Brun, Y.-Y. Chou, A. Lee, M. Barysheva, X. Pennec, K. McMahon, M. Meredith, G. De Zubicaray, M. Wright, A. W. Toga, and P. Thompson. Best individual template selection from deformation tensor minimization. In *Proc. of the 2008 IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'08), Paris, France, May 14-17*, pages 460–463, no address, France, 2008.
- [31] M. Lorenzi and X. Pennec. Geodesics, Parallel Transport & One-parameter Subgroups for Diffeomorphic Image Registration. *International Journal of Computer Vision*, 105(2):111–127, Nov. 2013.
- [32] J. S. Marron and A. M. Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753, September 2014.
- [33] K. Mcleod, M. Sermesant, P. Beerbaum, and X. Pennec. Spatio-Temporal Tensor Decomposition of a Polyaffine Motion Model for a Better Analysis of Pathological Left Ventricular Dynamics. *IEEE Transactions on Medical Imaging*, 34(7):1562–1675, July 2015.
- [34] M. Moakher. Means and averaging in the group of rotations. *SIAM Journal of Matrix Analysis and Applications*, 24(1):1–16, 2002.
- [35] J. Oller and J. Corcuera. Intrinsic analysis of statistical estimation. *Annals of Statistics*, 23(5):1562–1581, 1995.
- [36] X. Pennec. Probabilities and Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. In Cetin, A.E., Akarun, L., Ertuzun, A., Gurcan, M.N., Yardimci, and Y., editors, *Proc. of Nonlinear Signal and Image Processing (NSIP'99)*, volume 1, pages 194–198, June 20-23, Antalya, Turkey, Turkey, 1999. IEEE-EURASIP.
- [37] X. Pennec. Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006. A preliminary appeared as INRIA RR-5093, January 2004.
- [38] X. Pennec. Sample-limited L_p Barycentric Subspace Analysis on Constant Curvature Spaces. In *Geometric Sciences of Information (GSI 2017)*, volume 10589, pages 20–28, Paris, France, Nov. 2017. Springer International Publishing.
- [39] X. Pennec. Barycentric subspace analysis on manifolds. *Annals of Statistics*, 46(6A):2711–2746, 2018.
- [40] X. Pennec and V. Arsigny. Exponential Barycenters of the Canonical Cartan Connection and Invariant Means on Lie Groups. In F. Barbaresco, A. Mishra, and F. Nielsen, editors, *Matrix Information Geometry*, pages 123–168. Springer, May 2012.
- [41] X. Pennec, P. Fillard, and N. Ayache. A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, 66(1):41–66, 2006. A preliminary version appeared as INRIA Research Report 5255, July 2004.
- [42] X. Pennec, C. R. Guttmann, and J.-P. Thirion. Feature-based Registration of Medical Images: Estimation and Validation of the Pose Accuracy. In *Proc. of First Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, volume 1496 of LNCS, pages 1107–1114, Cambridge, USA, United States, 1998. Springer.

- [43] X. Pennec and M. Lorenzi. Beyond Riemannian Geometry The affine connection setting for transformation groups Chapter 5 of . In S. S. X. Pennec and T. Fletcher, editors, *Riemannian Geometric Statistics in Medical Image Analysis.*, chapter 5. Elsevier, Sept. 2019.
- [44] M.-M. Rohé, M. Sermesant, and X. Pennec. Barycentric Subspace Analysis: a new Symmetric Group-wise Paradigm for Cardiac Motion Tracking. In *MICCAI 2016 - the 19th International Conference on Medical Image Computing and Computer Assisted Intervention*, volume 9902 of *MICCAI 2016, Lecture Notes in Computer Science*, pages 300–307, Athens, Greece, Oct. 2016.
- [45] M.-M. Rohé, M. Sermesant, and X. Pennec. Low-Dimensional Representation of Cardiac Motion Using Barycentric Subspaces: a New Group-Wise Paradigm for Estimation, Analysis, and Reconstruction. *Medical Image Analysis*, 45:1–12, Apr. 2018.
- [46] C. Small. *The Statistical Theory of Shapes*. Springer series in statistics. Springer, 1996.
- [47] S. Sommer. Horizontal Dimensionality Reduction and Iterated Frame Bundle Development. In F. Nielsen and F. Barbaresco, editors, *Geometric Science of Information*, number 8085 in *Lecture Notes in Computer Science*, pages 76–83. Springer Berlin Heidelberg, 2013.
- [48] S. Sommer, F. Lauze, and M. Nielsen. Optimization over geodesics for exact principal geodesic analysis. *Advances in Computational Mathematics*, 40(2):283–313, June 2013.
- [49] C. Tobon-Gomez, M. De Craene, K. Mcleod, L. Tautz, W. Shi, A. Hennemuth, A. Prakosa, H. Wang, G. Carr-White, S. Kapetanakis, A. Lutz, V. Rasche, T. Schaeffter, C. Butakoff, O. Friman, T. Mansi, M. Sermesant, X. Zhuang, S. Ourselin, H. O. Peitgen, X. Pennec, R. Razavi, D. Rueckert, A. F. Frangi, and K. Rhode. Benchmarking framework for myocardial tracking and deformation algorithms: an open access database. *Medical Image Analysis*, 17(6):632–648, 2013.
- [50] G. S. Weyenberg. *Statistics in the Billera–Holmes–Vogtmann treespace*. PhD thesis, University of Kentucky, 2015.
- [51] L. Yang. Riemannian median and its estimation. *LMS Journal of Computation and Mathematics*, 13:461–479, 2010.
- [52] L. Yang. *Medians of probability measures in Riemannian manifolds and applications to radar target detection*. PhD thesis, Poitier University, December 2011.
- [53] H. Zhai. *Principal component analysis in phylogenetic tree space*. PhD thesis, University of North Carolina at Chapel Hill, 2016.
- [54] H. Ziezold. On Expected Figures and a Strong Law of Large Numbers for Random Elements in Quasi-Metric Spaces. In J. Kožešnik, editor, *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians*, number 7A, pages 591–602. Springer Netherlands, 1977.