Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Corso di Laurea Magistrale in Matematica

Barycentric Subspace Analysis on the Sphere and Image Manifolds

Relatore:
Chiar.ma Prof.
Giovanna Citti

Correlatore:
Chiar.mo Prof.
Xavier Pennec

Presentata da:
Sofia Farina

Sessione Unica
Anno Accademico 2016-2017

Introduction

The purpose of this thesis is to present a generalization of Principal Component Analysis (PCA) to Riemannian manifolds called Barycentric Subspace Analysis and show some applications.

In the Euclidean case the PCA is a statistical tool of multivariate analysis that has the purpose of reducing the dimension of the space in the presence of a large number of data, in fact it corresponds to search the subspace which maximizes the variance of the data. The limit of this method arises when we want to work on differentiable manifolds such as the Space of Images. For this reason, a generalization of the PCA becomes an interesting object of study. A definition that generalizes the concept of affine subspace to manifolds, i.e. the barycentric subspaces, was first introduced by X. Pennec in [7]. They lead to a hierarchy of properly embedded linear subspaces of increasing dimension, i.e. a flag of subspaces, which generalize the notion of affine Euclidean subspaces. In the Euclidean PCA the sum of the unexplained variance by all the subspaces of the flag is minimized, and a similar criterion applied to barycentric subspaces is at the basis of the Barycentric Subspace Analysis (BSA).

In this thesis we will present a detailed study of the method on the sphere since it can be considered as the finite dimensional projection of a set of probability densities that have many practical applications. This is the most original part of this thesis work. We also show an application of the barycentric subspace method for the study of cardiac motion in the problem of image registration, following the work of M.M. Rohé [13].

The thesis is organized as follows:

Chapter 1 recalls the main concepts of Differential Geometry, such as Riemannian manifolds, Levi Civita connections, and geodesics. In this regard, we show that if the metric is compatible with the connection, the geodesic minimizing the functional lengths are solutions of a differential equation that involves the connection. We will therefore show that the geodesics in the sphere are the great circle. In addition, we introduce the main notions of Groups and Lie algebras which we will use in Chapter 5 for the registration

of images.

To work with data on a manifold, we need to define the statistical tools on it. For this reason, in Chapter 2 the definitions of statistics on Riemannian manifolds are presented. In fact, the measure induced by the Riemannian metric on the manifold allows to define probability density functions. We focus on the definition of expected value of a random variable and we will also define the mean of Fréchet and Karcher defined as the locus of the points that locally or globally minimize the variance.

In Chapter 3 the definition of Barycentric Subspaces is given [7]. Depending on the definition of mean that we choose on the manifold: Fréchet, Karcher or exponential barycenter, we obtain three types of barycentric subspaces: Fréchet (FBS), Karcher (KBS) or exponential (EBS). It is shown that these three subspaces are related to each other since they are nested. In particular, the closure of the largest of these, EBS, leads to the definition of affine subspaces on the manifold. If the manifold is complete, the affine subspaces are also complete. Finally we use the barycentric subspaces for the generalization of the PCA to the Riemannian manifolds, since we define an increasing family of barycentric subspaces that generate a flag of subspaces. The chapter ends by showing three different numerical methods of analysis with barycentric subspaces.

In Chapter 4 we present an application of Barycentric Subspaces on the sphere, since the sphere can represent the space of probability density functions which can be used to model the frequency of pixel values in images. We first clarify that the space of probability density functions can be identified with a sphere in L^2, and we will restrict to a finite dimensional space in view of the implementation. Then we study the Riemannian geometry induced on the sphere and we show that the barycentric subspaces are the geodesics passing through the reference points. In conclusion, the three numerical methods described in Chapter 3 are tested on the sphere with different datasets.

Chapter 5 shows the use of Barycentric Subspaces in the image registration problem based on [13]. The registration of two images consists in matching the voxel intensities of the images and it is performed by minimizing a suitable energy functional depending on the distance between one image, and of the deformation of the second. Since the space of the images is not a linear space and can be described by a manifold, the distance and the functional will be expressed in terms of the exponential mapping. An algorithm called LCC Log Demons is presented which uses barycentric subspaces for cardiac image registration.

Introduzione

Lo scopo di questa tesi è di presentare una generalizzazione della Principal Component Analysis (PCA) alle varietà Riemanniane denominata Barycentric Subspace Analysis e mostrarne alcune applicazioni.

Nel caso Euclideo la PCA è uno strumento statistico di analisi multivariata che ha lo scopo di ridurre le dimensioni dello spazio in presenza di un numero elevato di dati, infatti corrisponde alla ricerca del sottospazio che massimizza la varianza dei dati. Il limite di questo metodo sorge quando si vuole lavorare in varietà differenziabili come ad esempio lo Spazio delle Immagini. Per questo motivo una generalizzazione della PCA diventa un interessante oggetto di studio. Richiameremo una definizione che generalizza il concetto di sottospazio affine alle varietà, introdotta per la prima volta da X. Pennec in [7], ovvero i sottospazi baricentrici. Essi generano una sequenza di sottospazi a bandiera (ovvero creano una gerarchia di sottospazi immersi che aumentano di dimensione) proprio come i sottospazi affine Euclidei. Nella PCA Euclidea viene minimizzata la somma della varianza di tutti i sottospazi della bandiera, applicando questo criterio ai sottospazi baricentrici viene definita la Barycentric Subspace Analysis (BSA).

In questa tesi verrà presentato uno studio completo del metodo sulla sfera perché può essere interpretata come la proiezione finito dimensionale di un'insieme di densità di probabilità che hanno moltissime applicazioni a livello pratico. Questa è la parte piu originale del presente lavoro di tesi. Viene inoltre mostrato un ulteriore applicazione del metodo dei sottospazi baricentrici per lo studio del moto cardiaco nel problema della registrazione delle immagini, seguendo il lavoro di M.M. Rohé [13].

La tesi è organizzata come segue:

Il Capitolo 1 richiama alcuni concetti fondamentali di Geometria Differenziale,come varietà Riemanniane, connessioni di Levi Civita, e geodetiche. A questo proposito mostreremo che le geodetiche minimizzanti il funzionale delle lunghezze sono caratterizzate come soluzioni di un'equazione differenziale che coinvolge la connessione, se la metrica è compatibile con la connessione. Mostreremo quindi che le geodetiche nella sfera sono le curve di

cerchio massimo. Inoltre nella parte conclusiva del capitolo introduciamo le nozioni principali di Gruppi e Algebre di Lie che ci serviranno nel Capitolo 5 per la registrazione di immagini.

Per lavorare con dati che vivono su una varietà bisogna definire gli strumenti statistici su essa. Per questo motivo nel Capitolo 2 vengono presentate le definizioni di statistica su varietà Riemanniane. Infatti la misura indotta dalla metrica Riemanniana sulla varietà permette di definire funzioni di densità di probabilità. Ci soffermiamo sulla definizione di valore atteso di una variabile random e definiremo inoltre la media di Fréchet e di Karcher definita come il luogo dei punti che minimizzano localmente o globalmente la varianza.

Nel Capitolo 3 viene data la definizione di Sottospazi Baricentrici [7], a seconda della definizione di media che scegliamo sulla varietà: Fréchet, Karcher o baricentro esponenziale, otteniamo tre tipi di sottospazi baricentrici: di Fréchet (FBS), di Karcher (KBS) o esponenziale (EBS). Viene quindi mostrato che questi tre sottospazi sono in relazione tra loro poichè contenuti uno nell'altro. In particolare la chiusura del più grande di essi, EBS, porta alla definizione di sottospazi baricentrici affini sulla varietà. Questi sottospazi affini se la varietà è completa sono essi stessi completi. Infine utilizziamo i sottospazi baricentrici per la generalizzazione della PCA alle varietà Riemanniane, poichè viene mostrato che generano una famiglia crescente di sottospazi formando una bandiera. Si conclude il capitolo mostrando tre diversi metodi numerici di analisi con i sottospazi baricentrici.

Nel Capitolo 4 viene mostrata un'applicazione della teoria dei sottospazi baricentrici alla sfera, questo poichè essa può rappresentare lo spazio delle funzioni di probabilità di densità (pdf) che viene utilizzato per modellizzare la frequenza dei valori dei pixel nelle immagini. Il Capitolo quindi introduce dapprima lo spazio delle pdf mostrando come esso porta allo studio della sfera. In seguito viene descritta la geometria Riemanniana indotta dall'ambiente Euclideo in cui la sfera è immersa. Mostreremo che in questo caso i sottospazi baricentrici sono le geodetiche passanti per i punti referenti. Infine vengono testati sulla sfera i tre metodi numerici descritti nel Capitolo 3 con diversi dataset.

Nel Capitolo 5 viene mostrato un utilizzo dei sottospazi baricentrici nel problema della registrazione delle immagini basato su [13]. La registrazione tra due immagini consiste nel trovare le corrispondenze nell'intensità dei voxel delle immagini e si trova minimizzando un funzionale di tipo energia che dipende dalla distanza tra le immagini, e dalla deformazione della seconda immagine. Poichè lo spazio delle immagini è uno spazio non lineare, può essere espresso da una varietà, la distanza e il funzionale saranno quindi espressi in termini della mappa esponenziale. Viene quindi presentato un algoritmo

chiamato LCC Log Demons che viene usato con i sottospazi baricentrici per la registrazione di immagini cardiache.

Contents

Chapter 1

Riemannian Geometry

The aim of this chapter is to introduce the main geometrical instruments we need.
The first part recall the notion of Riemannian geometry, and the concept of geodesics, which it is the shortest path connecting two points on the manifold. The main concepts are taken from [9], [16] and [19].
The second part introduces the theory of Lie groups and Lie algebra focusing on the one-parameter subgroups. The references are taken from [9], [18] and [16].

1.1 Smooth Manifolds

A topological space $\mathcal{M}$ is said a *topological Manifold* of dimension n if it is a Hausdorff, second countable and locally an Euclidean space, i.e. every point of $\mathcal{M}$ has an open neighborhood homeomorphic to an open set in n-dimensional Euclidean space $\mathbb{R}^n$.

A *chart* is a couple (U, Φ), where U is a open subset of the manifold $\mathcal{M}$, and $\Phi : U \to \mathbb{R}^n$ a homeomorphism.
Two charts (u, Φ) and (v, Ψ) of a topological manifold are C^∞-*compatible* if the two maps

$$\Phi \circ \Psi^{-1} : \Psi(U \cap V) \to \Phi(U \cap V), \quad \Psi \circ \Phi^{-1} : \Phi(U \cap V) \to \Psi(U \cap V)$$

are C^∞.

A C^∞-*atlas* is a collection $\mathfrak{A} = \{(U_\alpha, \Phi_\alpha)\}$ of pairwise C^∞-compatible charts that cover the manifold $\mathcal{M}$, i.e. such that $M = \bigcup_\alpha U_\alpha$.
An atlas $\mathfrak{A}$ on a locally Euclidean space is said to be *maximal* if it is not contained in a larger atlas.

Definition 1.1. A *smooth* or C^∞ *manifold* is a topological space $\mathcal{M}$ together with a maximal atlas. The maximal atlas is also called a *differentiable structure* on $\mathcal{M}$.

1.1.1 Submanifolds embedded in $\mathbb{R}^n$

Let $\mathcal{M}$ be an n-dimensional manifold, a $k \leq n$ submanifold of M is a subset $S \subset \mathcal{M}$ such that for every point $p \in S$ there exists a chart (U, Φ) containing p such that $\Phi(S \cap U)$ is the intersection of a k-dimensional plane with $\Phi(U)$. The pairs $(S \cap U, \Phi|_{S\cap U})$ form an atlas for the differential structure on S.

Proposition 1.2. *Let $\mathcal{M}$ be a subset of Euclidean space $\mathbb{R}^n$. Then the following are equivalent:*

1. *$\mathcal{M}$ is a k-dimensional submanifold;*

2. *$\mathcal{M}$ is a k-dimensional manifold, and can be given a differentiable structure in such a way that the inclusion $i : \mathcal{M} \hookrightarrow \mathbb{R}^n$ is an embedding;*

3. *For every $x \in \mathcal{M}$ there exists an open set $V \subseteq \mathbb{R}^n$ containing x and an open set $W \subseteq \mathbb{R}^n$ and a diffeomorphism $F : V \to W$ such that $F(\mathcal{M} \cap V) = (\mathbb{R} \times \{0\}) \cap W$;*

4. *$\mathcal{M}$ is locally the zero set of a submersion: for every $x \in \mathcal{M}$ there exists an open set V containing x and a submersion $\Phi : V \to Z \subseteq \mathbb{R}^{n-k}$ such that $\mathcal{M} \cap V = \Phi^{-1}(0)$.*

Theorem 1.3. *[Inverse function Theorem] Let F be a smooth function from an open neighbourhood of $x \in \mathbb{R}^n$ to $\mathbb{R}^n$, such that the derivative D_xF is an isomorphism. Then there exists an open set A containing x and an open set B containing $F(x)$ such that $F|_A$ is a diffeomorfism from A to B.*

The proposition implies that:

1. The sphere $S^n = \{x \in \mathbb{R}^{n+1} : |x| = 1\}$ is a submanifold of $\mathbb{R}^n$

2. The orthogonal group $O(n)$ of the $n \times n$ matrices A satisfying $A^TA = Id$ is a submanifold of $Gl_n(\mathbb{R})$

3. The special orthogonal group $SO(n)$ of the $n \times n$ matrices A satisfying $A^TA = Id$ and $\det A = 1$ is a submanifold of $Gl_n(\mathbb{R})$

1.1.2 Tangent space

Let $\mathcal{M}$ a manifold and $p \in \mathcal{M}$, it is possible to define the notion of tangent space at this point.

Definition 1.4. Let $p \in \mathcal{M}$ be any point of the n-dimensional manifold $\mathcal{M}$, given two C^1-curves $\gamma_1 :]-\epsilon, \epsilon[\to \mathcal{M}$ and $\gamma_2 :]-\epsilon, \epsilon[\to \mathcal{M}$ passing through p (i.e. $\gamma_1(0) = \gamma_2(0) = p$) are *equivalent* if and only if there is some chart (U, Φ) at p such that:

$$(\Phi \circ \gamma_1)'(0) = (\Phi \circ \gamma_2)'(0)$$

Definition 1.5 (Tangent Vectors)**.** Given a manifold $\mathcal{M}$ for any $p \in \mathcal{M}$ a *tangent vector to* $\mathcal{M}$ at p is any equivalence class of C^1-curves through p on $\mathcal{M}$, modulo the equivalence relation defined on definition 1.4. The set of all tangent vectors at p is denoted by $T_p(\mathcal{M})$:

$$T_p\mathcal{M} = \{\gamma : (-\epsilon, \epsilon) \to \mathcal{M} \quad \text{s.t.}\, \gamma(0) = p\} \,/\sim$$

Definition 1.6. A smooth *vector field* X on a manifold $\mathcal{M}$ is a linear map $X : C^\infty(\mathcal{M}) \to C^\infty(\mathcal{M})$ such that:

$$X(fg) = fX(g) + gX(f) \quad \forall f, g \in C^\infty(\mathcal{M})$$

1.1.3 Affine Connection on Smooth Manifolds

The main obstacle to the definition of differential of a vector field X, is that for every point p, q, $X(p)$ and $X(q)$ belong to the tangent space to the manifold $\mathcal{M}$ at different points and can not be subtracted. It is necessary to introduce the notion of affine connection.

Definition 1.7. Let $\mathcal{M}$ be a smooth manifold, an *affine connection* on $\mathcal{M}$ is a differential operator, send smooth vector fields X and Y to a smooth vector field $\nabla_X Y$:

$$\begin{aligned} \nabla : \mathcal{C}^\infty(\mathcal{M}, T\mathcal{M}) \times \mathcal{C}^\infty(\mathcal{M}, T\mathcal{M}) &\to \mathcal{C}^\infty(\mathcal{M}, T\mathcal{M}) \\ (X, Y) &\mapsto \nabla_X Y \end{aligned}$$

which satisfies for all smooth vector fields X, Y and Z and real-vaued functions f on $\mathcal{M}$:

1. $\nabla_{X+Y} Z = \nabla_X Z + \nabla_Y Z$

2. $\nabla_{fX}Y = f\nabla_X Y$
3. $\nabla_X(Y+Z) = \nabla_X Y + \nabla_X Z$
4. $\nabla_X(fY) = X[f]Y + f\nabla_X Y$

The vector field $\nabla_X Y$ is known as the *covariant derivative* of the vector field Y along X (with respect to the connection). The *torsion tensor* T of an affine connection ∇ is the operators sending smooth vector fields X and Y on $\mathcal{M}$ to the smooth vector fields $T(X,Y)$ given by:

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]$$

where $[X,Y](f) = XY(f) - YX(f)$ for all f real-valued function.
If the manifold is equipped with an affine connection, then this connection allows one to transport vectors of the manifold along curves so that they stay parallel with respect to the connection, this idea is at the basis of the definition of parallel transport which is about to be introduced.

Remark 1.8. An affine connection ∇ on $\mathcal{M}$ is said to be *torsion-free* if its torsion tensor is everywhere zero, so that $\nabla_X Y - \nabla_Y X = [X,Y]$ for all smooth vector fields X and Y on $\mathcal{M}$

Definition 1.9. If $\gamma : [a,b] \to \mathcal{M}$ is a smooth curve and $\xi \in T_x\mathcal{M}$, where $x = \gamma(a)$, then a vector field X along γ (and in particular, the value of this vector field at $y = \gamma(b)$) is called the *parallel transport of* ξ *along* γ if

1. $\nabla_{\gamma(t)} X = 0 \; \forall\, t \in [a,b]$
2. $X_{\gamma(a)} = \xi$

It is necessary to define the covariant derivative, before giving the definition of geodesic:

Definition 1.10. A smooth vector field along the curve $\gamma : I \to \mathcal{M}$ is a smooth map $V : I \to T\mathcal{M}$ such that $V(t) \in T_{\gamma(t)}\mathcal{M}$ for all $t \in I$. It is called $\mathcal{T}(\gamma)$ the space of the smooth vector field along γ.
A connection ∇ on the manifold $\mathcal{M}$ defines an unique operator for every curves $\gamma : I \to \mathcal{M}$:

$$D_t : \mathcal{T}(\gamma) \to \mathcal{T}(\gamma)$$

such that:

1. D_t is linear on $\mathbb{R}$: $D_t(aV + bW) = aD_tV + bD_tW \quad \forall\, a,b \in \mathbb{R}$
2. D_t satisfy: $D_t(fV) = \dot{f}V + fD_tV \quad \forall\, f \in C^\infty(I)$

Definition 1.11. Let ∇ be a connection on a manifold $\mathcal{M}$ and let $\gamma : I \to \mathcal{M}$ a curve on $\mathcal{M}$. It is said that γ s a *geodesic* for ∇ if and only if

$$\mathrm{D}_t\dot{\gamma} \equiv 0$$

1.2 Riemannian Manifolds

Definition 1.12. A *Riemannian Metric* on a differentiable manifold $\mathcal{M}$ is given by a scalar product on each tangent space $T_p\mathcal{M}$ which depends C^∞ on the base point $p \in \mathcal{M}$. A *Riemannian manifold* is a differentiable manifold, equipped with a Riemannian metric.

In local coordinates, a metric is represented by a positive definite, symmetric matrix called *local representation of the Riemannnian metric* in the chart x:

$$(g_{ij}(x))_{i,j=1,\cdots,n}$$

defined on $\mathcal{M}$.

The product of two tangent vectors $v, w \in T_p\mathcal{M}$ with coordinate representations $(v^1, ..., v^n)$ and $(w^1, ..., w^n)$ respectively, is:

$$< v, w >_g = g_{ij}(x(p))v^i w^j$$

In particular, $< \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} >= g_{ij}$, where $\frac{\partial}{\partial x^i}$ is a base of the tangent space $T_p\mathcal{M}$. The length of of v is given by $\|v\| =< v, v >^{\frac{1}{2}}$.

Theorem 1.13. *Each differentiable manifold may be equipped with a Riemannian metric.*

Proof. see [4] □

1.2.1 Riemmanian distance and geodesics

Given a regular curve γ, the norm of its derivative is well defined and allows to define the length of the curve:

Definition 1.14. Let $(\mathcal{M}, g)$ be a connected manifold. The *length* on the metric g of a smooth curve $\gamma : [a, b] \to \mathcal{M}$ is the following:

$$L(\gamma) = \int_a^b \left\|\dot{\gamma(t)}\right\|_g \mathrm{d}t = \int_a^b \left(< \dot{\gamma(t)}, \dot{\gamma(t)} >_{\gamma(t)}\right)^{\frac{1}{2}} \mathrm{d}t$$

In local coordinate $x = (x^1(\gamma(t)), ..., x^n(\gamma(t)))$:

$$L(\gamma) = \int_a^b \sqrt{g_{ij}(x(\gamma(t))\frac{\mathrm{d}x^i}{\mathrm{d}t}(\gamma(t))\frac{\mathrm{d}x^j}{\mathrm{d}t}(\gamma(t))}\mathrm{d}t$$

Definition 1.15. Let $\mathcal{M}$ be a connected manifold and $p, q \in \mathcal{M}$. Let $\langle ., . \rangle_g$ be a Riemannian metric on $\mathcal{M}$. The *distance* between two points p and q is defined as:

$$d_g(p,q) := \inf \{ L(\gamma) \,|\, \gamma : [a,b] \to \mathcal{M} \text{ piecewise smooth curves with } \gamma(a) = p, \gamma(b) = q \}$$

Lemma 1.16. *The distance function satisfies the axioms:*

1. $d_g(p,q) \geq 0 \quad \forall p, q \in \mathcal{M}$;
   $d_g(p,q) = 0 \Leftrightarrow p = q$;

2. $d_g(p,q) = d_g(q,p)$;

3. $d_g(p,q) \leq d_g(p,r) + d_g(r,q) \quad \forall p, q, r \in \mathcal{M}$.

Proof. The proof is shown in [4]. □

Definition 1.17. The curves realizing the minimum of the Riemannian distance for any two points of the manifold are called *minimizing geodesics.*

Definition 1.18. The manifold is said *geodesically complete* if any geodesic can be defined on the whole $\mathbb{R}$:

$$\gamma : \mathbb{R} \to \mathcal{M}$$

If a manifold is geodesically complete, it has no boundary nor any singular point that can be reached in a finite time.

Theorem 1.19 (Hopf-Rinow-De Rham theorem)**.** *Let $\mathcal{M}$ be a connected Riemannian manifold. Then the following statements are equivalent:*

- *$\mathcal{M}$ is a complete metric space*
- *The closed and bounded subset of $\mathcal{M}$ are compact*
- *$\mathcal{M}$ is geodesically complete*

Furthermore, this theorem implies that given any two points p and q in $\mathcal{M}$, there exists a length minimizing geodesic connecting these two points.

1.3 The Levi-Civita Connection

Let $(\mathcal{M}, g)$ be a Riemannian manifold and let ∇ be an affine connection on $\mathcal{M}$. It is said that ∇ is *compatible with the Riemannian metric* g if:

$$Z[g(X,Y)] = g(\nabla_Z X, Y) + g(X, \nabla_Z Y)$$

for all smooth vector fields X, Y and Z on $\mathcal{M}$. On every Riemannian manifold there exists a unique torsion-free connection that is compatible with the Riemannian metric.

Theorem 1.20. *Let $(\mathcal{M}, g)$ be a Riemannian manifold. Then there exists a unique torsion-free affine connection ∇ on $\mathcal{M}$ compatible with the Riemannian metric g. This connection is characterized by the identity:*

$$\begin{aligned} 2g(\nabla_X Y, Z) =& X[g(Y,Z)] + Y[g(X,Z)] - Z[g(X,Y)] + \\ &+ g([X,Y]Z) - g([X,Z], y) - g([Y,Z], X) \end{aligned}$$

for all smooth vector fields X,Y and Z on $\mathcal{M}$.

Proof. The proof is shown in [19] □

It is known as *Levi-Civita connection*. The geodesics computed with respect to this connection are called *Riemannian geodesics*.

Definition 1.21. Let $(\mathcal{M}, g)$ be a Riemannian manifold of dimension n, and let $\gamma : I \to \mathcal{M}$ be a smooth curve in $\mathcal{M}$, defined over some interval I in $\mathbb{R}$. γ is a *Riemannian geodesic* if and only if

$$D_t\left(\frac{d\gamma(t)}{dt}\right) = 0 \tag{1.1}$$

Definition 1.22. Given a chart $(x_1, \cdots, x_n)$ over some open set $U \in \mathcal{M}$, the *Christoffel symbols* are defined:

$$\Gamma^i_{jk} = \frac{1}{2} g^{im}(\partial_k g_{mj} + \partial_j g_{mk} - \partial_m g_{jk})$$

where $(g^{ij}) = (g_{ij})^{-1}$ is the inverse of the metric matrix.

The Levi-Civita connection is determined in a local system through the Christoffel symbols:

$$\nabla_{\partial_j} \partial_k = \sum_{i=1}^{n} \Gamma^i_{jk} \partial_i \tag{1.2}$$

then $\Gamma^i_{jk} = \Gamma^i_{kj}$.

Let $\gamma : I \to U$ be a smooth curve in U, and let $\gamma^i(t) = x^i \circ \gamma(t)$ for all $t \in \gamma^{-1}(U)$.Then

$$\frac{\mathrm{d}\gamma(t)}{\mathrm{d}t} = \sum_{k=1}^{n} \frac{\mathrm{d}\gamma^k(t)}{\mathrm{d}t}\partial_k$$

So that:

$$\begin{aligned}\frac{D}{\mathrm{d}t}\frac{\mathrm{d}\gamma(t)}{\mathrm{d}t} &= \sum_{k=1}^{n}\left(\frac{\mathrm{d}^2\gamma^k(t)}{\mathrm{d}t^2}\partial_k + \frac{\mathrm{d}\gamma^k(t)}{\mathrm{d}t}\sum_{j=1}^{n}\frac{\mathrm{d}\gamma^j(t)}{\mathrm{d}t}\nabla_{\partial_j}\partial_k\right)\\ &= \sum_{i=1}^{n}\left(\frac{\mathrm{d}^2\gamma^i(t)}{\mathrm{d}t^2} + \sum_{j=1}^{n}\sum_{k=1}^{n}\Gamma^i_{jk}(\gamma(t))\frac{\mathrm{d}\gamma^j(t)}{\mathrm{d}t}\frac{\mathrm{d}\gamma^k(t)}{\mathrm{d}t}\right)\partial_i\end{aligned}$$

So $\gamma : I \to U$ is a geodesic if and only if:

$$\frac{\mathrm{d}^2\gamma^i(t)}{\mathrm{d}t^2} + \sum_{j=1}^{n}\sum_{k=1}^{n}\Gamma^i_{jk}(\gamma(t))\frac{\mathrm{d}\gamma^j(t)}{\mathrm{d}t}\frac{\mathrm{d}\gamma^k(t)}{\mathrm{d}t} = 0 \quad i = 1, \cdots, n \tag{1.3}$$

Next theorem ensures that geodesics defined by means of the connection, are indeed minimizing geodesics in the sense of definition 1.17 above if and only if the connection is compatible with the Riemannian metric.

Theorem 1.23. *Let p and q be distinct points in a Riemannian manifold $(\mathcal{M}, g)$ and let $\gamma : [a, b] \to \mathcal{M}$ be a piecewise smooth curve in $\mathcal{M}$ from p to q, parametrized by the arclength. Suppose that the length of γ is less than or equal to the length of every other piecewise smooth curve from p to q. Then γ is a smooth geodesic in $\mathcal{M}$.*

Proof. The proof of this theorem can be viewed in [19]. □

It is shown that the geodesics in the sphere S^n are the great circle.

Example 1.24 (Sphere S^2). The Riemannian metric on the sphere is induced by the embedded space $\mathbb{R}^3$, indeed the more simple Riemannian metric is the Euclidean metric, which metric matrix is the identity $g = Id$ or can also written in Cartesian coordinates by $g = dx^2 + dy^2 + dz^2$.
It is chosen spherical coordinates on the sphere $S^2 \subset \mathbb{R}^3$:

$$\begin{cases} x = R\sin\phi\cos\theta \\ y = R\sin\phi\sin\theta \\ z = R\cos\phi \end{cases} \tag{1.4}$$

where $\theta \in (-\pi, \pi)$ and $\phi \in (0, \pi)$, while R is the radius of the sphere. Computing dx,dy and dz in spherical coordinates, it is found:

$$\begin{cases} dx = R(\phi' \cos\phi \cos\theta - \theta' \sin\phi \sin\theta) \\ dy = R(\phi' \cos\phi \sin\theta + \theta' \sin\phi \cos\theta) \\ dz = R\phi' \sin\phi \end{cases} \tag{1.5}$$

Hence, the metric on the sphere is:

$$g_{sphere} = R^2 d\phi^2 + R^2 \sin^2\phi d\theta^2$$

So the metric matrix is:

$$g_{sphere} = \begin{pmatrix} R^2 & 0 \\ 0 & R^2 \sin^2\phi \end{pmatrix} \tag{1.6}$$

Formula 1.3 can be applied to compute geodesics. So that the computation of the Christoffel symbols is necessary. By definition 1.22:

$$\begin{aligned} \Gamma^{\theta}_{\phi\theta} &= \frac{(\partial_\phi g_{\theta\phi} + \partial_\theta g_{\phi\phi} - \partial_\phi g_{\phi\theta}) g^{\phi\theta}}{2} + \frac{(\partial_\phi g_{\theta\theta} + \partial_\theta g_{\phi\theta} - \partial_\theta g_{\phi\theta}) g^{\theta\theta}}{2} \\ &= \partial_\phi (R^2 \sin\phi) \frac{1}{2R^2 \sin^2\phi} = \frac{\cos\phi}{\sin\phi} = \arctan\phi \end{aligned}$$

$$\begin{aligned} \Gamma^{\phi}_{\theta\theta} &= \frac{(\partial_\theta g_{\theta\phi} + \partial_\theta g_{\theta\phi} - \partial_\phi g_{\theta\theta}) g^{\phi\phi}}{2} + \frac{(\partial_\theta g_{\theta\theta} + \partial_\theta g_{\theta\theta} - \partial_\theta g_{\theta\theta}) g^{\theta\phi}}{2} \\ &= \partial_\phi (R^2 \sin\phi) \frac{1}{2R^2} = -\sin\phi\cos\phi \end{aligned}$$

$$\Gamma^{\theta}_{\theta\theta} = \frac{(\partial_\theta g_{\theta\theta}) g^{\theta\theta}}{2} = 0$$

The other Christoffel symbols are zero $\Gamma^{\theta}_{\theta\theta} = \Gamma^{\theta}_{\phi\phi} = \Gamma^{\phi}_{\theta\theta} = \Gamma^{\phi}_{\phi\phi} = 0$.
Then, the geodesics are the curves $\gamma(t) = (\theta(t), \phi(t))$ that satisfy that differential equation:

$$\begin{cases} \ddot{\phi} - \dot{\theta}^2 \sin\phi\cos\phi = 0 \\ \ddot{\theta} + \dot{\phi}\dot{\theta} \arctan\phi = 0 \end{cases} \tag{1.7}$$

Since the meridians are defined by $\ddot{\phi} = \ddot{\theta} = \dot{\theta} = 0$, and they satisfy the system above, it can be claimed that the meridians are geodesics.

Proposition 1.25. *The geodesics on a sphere are the arcs from great circles, that is, arcs from circles formed by the intersection of a plane containing the center of the sphere with the sphere itself.*

1.3.1 Exponential and Logarithmic map

Formula 1.3 and standard existence and uniqueness theorems for solutions of ordinary differential systems of equations ensure that, given a tangent vector $\partial_v \in T_x\mathcal{M}$ at any point $x \in \mathcal{M}$ and given any real number t_0, exists a unique maximal geodesic $\gamma : I \to \mathcal{M}$ defined on some open interval I containing t_0, such that $\gamma(t_0) = x$ and $\gamma'(t_o) = \partial_v$. Moreover exists $\delta_x > 0$ such that it can be defined the *exponential map* as:

$$exp_x : \{\partial_v \in T_x\mathcal{M} : |\partial_v| < \delta_x\} \to \mathcal{M}$$

at x is defined by $exp_x(\partial_v) = \gamma_{(x,\partial_v)}(1)$.

A corollary of the Hopf-Rinow De Rham Theorem 1.19 can be stated as follows:

Corollary 1.26. *Let $(\mathcal{M}, g)$ be a connected Riemannian manifold. Then the following three conditions are equivalent:*

1. *the Riemannian distance function on $\mathcal{M}$ is complete (i.e. every Cauchy sequence in $\mathcal{M}$ converges);*
2. *the Riemannian manifold $(\mathcal{M}, g)$ is geodesically complete;*
3. *there exists a point p on $\mathcal{M}$ with the property that the exponential map exp_p is defined over the whole tangent space $T_p\mathcal{M}$ to $\mathcal{M}$ at p (i.e. every geodesic passing through the point p can be extended to a geodesic from $\mathbb{R}$ into $\mathcal{M}$)*

Then the *exponential map* is defined in the whole tangent space:

$$exp_x : T_x\mathcal{M} \to \mathcal{M} \tag{1.8}$$

$$\partial_v \mapsto exp_x(\partial_v) = \gamma_{(x,\partial_v)}(1) \tag{1.9}$$

The exponential is a local diffeomorphism and its inverse is the *logarithmic map* denoted by $\vec{pq} = Log_p(q)$, this is the smallest vector (in norm) such that $q = exp_p(\vec{pq})$.

$$Log_p : \mathcal{M} \to T_p\mathcal{M}$$

$$q \mapsto \vec{pq} = Log_p(q)$$

In this chart the geodesics going through p are represented by the lines going through the origin $Log_p\gamma_{(p,\vec{pq})}(t) = t\,\vec{pq}$.

Since the exponential map is a local diffeomorphism, the distance defined on the manifold is locally equivalent to a norm defined on the tangent plane. Indeed there exist c_1, c_2 constants such that, for all $q = exp_p(v)$, satisfy:

$$c_1\|v\| \leq \mathrm{d}(p,q) \leq c_2\|v\|$$

Example 1.27 (Sphere). Given two points on the sphere p and $q \in S^n$, which are parametrized by the vectors starting from the origin o: $p = \vec{op}$ $q = \vec{oq}$. The distance induced from the metric is $d(p,q) = \arccos(p^T q) = \theta \in [0,\pi]$.
Given $x \in S^n$ and $v \in T_x S^n$ such that $\|v\| = 1$. The curve $\gamma_v(t) = \cos(t)x + \sin(t)v$ is the only geodesic which satisfy $\gamma_v(0) = x$ and $\gamma_v'(0) = v$. Then from the definition the exponential map:

$$\begin{aligned} \exp_x : T_x S^n &\to S^n \\ v &\to \gamma_v(1) \end{aligned}$$

Therefore $\exp_x(v) = \cos(|v|)x + \sin(|v|)v$.
The inverse is the Logarithmic map:

$$\begin{aligned} Log_x : S^n &\to T_x S^n \\ y &\to \frac{\theta}{\sin\theta}(y - \cos(\theta)x) \end{aligned}$$

1.3.2 Cut Locus

Definition 1.28 (Cut locus and Tangential cut locus)**.** If $\gamma_{(x,\partial_v)}(t) = exp_x(t\partial_v)$ is a geodesic defined on $[0, +\infty[$, it can be minimizing only up to a time $t_0 < +\infty$ and not minimizing for any instant of time $t > t_0$. In this case the points $z = \gamma_{(x,\partial_v)}(t_0)$ is called *cut point* and $t_0\partial_v$ a *tangential cut point*. The set of all cut points of all geodesics starting from x is the *cut locus* $C(x) \in \mathcal{M}$ and the set of corresponding vectors *tangential cut locus* $\mathcal{C}(x) \in T_x\mathcal{M}$.

Example 1.29. (sphere)

[H]

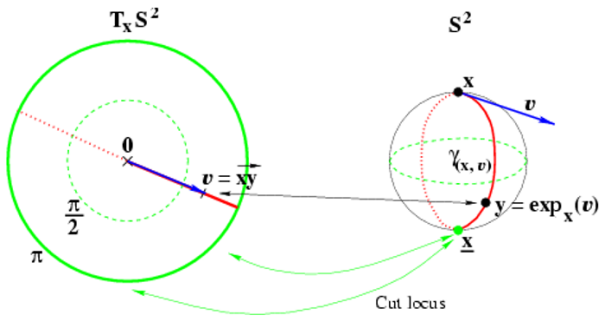


Figure 1.1: This image is taken from [10]

In the sphere the cut locus of a point p is its antipodal point $-p$

1.4 Lie Groups

Definition 1.30. A *Lie Group* is a smooth manifold G which is also a group such that the two group operations, multiplication and inverse, are smooth maps.

$$\begin{aligned} m : G \times G \to G \in \mathcal{C}^\infty \\ (g,h) \to gh \end{aligned} \qquad \begin{aligned} i : G \to G \in \mathcal{C}^\infty \\ g \to g^{-1} \end{aligned}$$

Definition 1.31. A map $F : H \to G$ between two Lie groups H and G is a *Lie group homonorphism* if it is a smooth map and a group homomorphism.

The group of homomorphism condition means that for all $h, x \in H$: $F(hx) = F(h)F(x)$. This can be written using the *left multiplication* which is the diffeomorphism $\ell_a : G \to G$ for an element a in a Lie group G, in the following way: $F \circ \ell_h = \ell_{F(h)} \circ F$ for all $h \in H$.
The general definition of Lie algebra is given by:

Definition 1.32. Let K be a field. A vector space $\mathfrak{g}$ over K is called *Lie Algebra* if there exists:

$$[\,,\,] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$$

with the following properties $X, Y, Z \in \mathfrak{g}$:

- $[X,Y] = -[Y,X]$
- Jacobi identity: $[X\,[Y,Z]] + [Y\,[Z,X]] + [Z,[X,Y]]$

Since the left translation of a Lie group G by an element $g \in G$ is a diffeomorphism that maps a neighborhood of the identity to a neighborhood of g, all the local information about the group is concentrated in a neighborhood of the identity.
Moreover, on the tangent space T_eG can be given a Lie bracket $[,]$, so it becomes a Lie algebra of the Lie group.
The Lie bracket on the tangent space T_eG is defined using the canonical isomorphism between the tangent space at the identity and the vector space of the left invariant vector fields on G:

Definition 1.33. Let G Lie Group and $X \in X(G)$ a vector field, X is said *left invariant* if

$$\mathrm{d}_{\ell_g} X = X \circ \ell_g$$

where $\ell_g : G \to G\, h \mapsto gh$.

The *Lie Algebra of the Lie Group* is defined as:

$$\mathrm{Lie}(G) = \mathfrak{g} = \left\{X \in X(G) | \mathrm{d}_{\ell_g} X = X \circ \ell_g\right\}$$

Theorem 1.34 (Von Neumann Theorem)**.** *Let G be a real algebraic group. Then G is a closed real Lie subgroup of $Gl(n, \mathcal{R})$*

Example 1.35. This Theorem implies that:

1. $Sl_n(\mathbb{R})$
2. $SO(n)$
3. $O(n)$

are Lie subgroups in $Gl_n\mathbb{R}$

Theorem 1.36. *There is a vector space isomorphism between:*

$$\begin{gathered} Lie\ (G) \leftrightarrow T_e(G) \\ X \mapsto X_e \\ \tilde{A} \nleftarrow A \end{gathered}$$

Proof. A complete proof is shown in [16]. □

Using this theorem, in groups we can define a group exponential map, analogous to the one defined in 1.8 , but with values on the algebra instead of the tangent plane.

1.4.1 Group Exponential map

Firstly, we need the following theorem to define the one-parameter subgroups:

Theorem 1.37. *Let G and H be two Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$ respectively and G simply connected. Let $\Psi : \mathfrak{g} \to \mathfrak{h}$ be a homomorphism. Then there exists a unique homomorphism $\Phi : G \to H$ such that $d\Phi = \Psi$.*

Proof. In [18]. □

Definition 1.38. Let G be a Lie Group a *one-parameter subgroup* is a homomorphism of Lie group $\alpha : \mathbb{R} \to G$

Let G be a Lie group and $\mathfrak{g}$ its Lie algebra. Let $X \in \mathfrak{g}$ then:

$$\lambda \frac{d}{dr} \mapsto \lambda X$$

is an homomorphism of the Lie algebra of $\mathbb{R}$ into $\mathfrak{g}$. Since the real line is simply connected, by Theorem 1.37 there exists a unique one parameter subgroup:

$$\exp_X : \mathbb{R} \to \mathfrak{g}$$

In other words, $t \mapsto \exp_X(t)$ is the unique one-parameter subgroup of G whose tangent vector at 0 is $X(e)$. This allows to define the group exponential:

Definition 1.39. Let G be a Lie Group, we call *group exponential of* G the map:

$$\exp : \mathfrak{g} \to G$$

setting $\exp(X) = \exp_X(1)$.

The following proposition, analogous to 1.26, but with domain on the algebra instead of the tangent space holds.

Proposition 1.40. *Let X belong to the Lie algebra $\mathfrak{g}$ of the Lie group G. Then*

1. $\exp(tX) = \exp_X(t)$ *for each* $t \in \mathbb{R}$
2. $\exp(t+s)X = (\exp tX)(\exp sX)$ *for all* $t, s \in \mathbb{R}$
3. $\exp(-tX) = (\exp tX)^{-1}$ *for each* $t \in \mathbb{R}$
4. $\exp : \mathfrak{g} \to G$ *is C^∞ and* $d\exp : \mathfrak{g}_0 \to G_e$ *is the identity map, so* exp *gives a diffeomorphism of a neighborhood of* 0 *in* $\mathfrak{g}$ *onto a neighborhood of* e *in* G.
5. $\ell_\sigma \circ \exp_X$ *is the unique integral curve of* X *which takes the value* σ *at* 0. *As a particular consequence, left invariant vector fields are always complete.*
6. *the one-parameter group of diffeomorphisms* X_t *associated with the left invariant vector field* X *is given by* $X_t = \alpha_{\exp_X(t)}$

Proof. The proof can be found in [18]. □

In particular if the bracket of two vector fields is zero, there is the following consequence:

Proposition 1.41. *Let G be a Lie group and $\mathfrak{g}$ its Lie algebra. If $X, Y \in \mathfrak{g}$ and $[X, Y] = 0$ then* $\exp X \exp Y = \exp(X + Y)$

1.4.2 Baker-Campbell-Hausdorff Formula

In general is not true that $\exp(X+Y) = \exp(X)\exp(Y)$, but the $\exp(X+Y)$ can be computed via the Baker-Campbell-Hausdorff formula. This formula links Lie groups to Lie algebras by expressing the logarithm of the product of two Lie group elements as a Lie algebra element using only Lie algebraic operations.
The Baker-Campbell-Hausdorff formula implies that if $X, Y \in \mathfrak{g}$ then

$$Z = \log(\exp(X)\exp(Y))$$

can formally be written as infinite sum of elements of $\mathfrak{g}$. This infinite series may or may not converge, so it need not define an actual element $Z \in \mathfrak{g}$.
The following general combinatorial formula was introduced by Dynkin:

$$\log(\exp X \exp Y) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \sum_{r_j+s_j>0} \frac{[X^{r_1}Y^{s_1}X^{r_2}Y^{s_2}\ldots X^{r_n}Y^{s_n}]}{\sum_{i=1}^{n}(r_i+s_i)\prod_{i=1}^{n} r_i! s_i!}$$

where the sum is performed over all nonegative values of s_i and r_i, and the following notation has been used:

$$[X^{r_1}Y^{s_1}X^{r_2}Y^{s_2}\cdots X^{r_n}Y^{s_n}] = [\underbrace{X,[X,\cdots[X}_{r_1},\underbrace{Y,[Y,\cdots[Y}_{s_1},\cdots\underbrace{X,[X,\cdots[X}_{r_n},\underbrace{Y,[Y,\cdots Y]}_{s_n}]\cdots]]$$

Since $[A,A] = 0$ the term is zero if $s_n > 1$ or if $s_n = 0$ and $r_n > 1$.
The firs few terms are well-known:

$$\begin{aligned}
Z(X,Y) =& \log(\exp X \exp Y) \\
=& X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}([X,[X,Y]] + [Y,[Y,X]]) \\
&- \frac{1}{24}[Y,[X,[X,Y]]] \\
&- \frac{1}{720}([Y,[Y,[Y,[Y,X]]]] + [X,[X,[X,[X,Y]]]]) \\
&+ \frac{1}{360}([X,[Y,[Y,[Y,X]]]] + [Y,[X,[X,[X,Y]]]]) \\
&+ \frac{1}{120}([Y,[X,[Y,[X+Y]]]] + [X,[Y,[X,[Y,X]]]]) + O((X,Y)^6)
\end{aligned}$$

1.4.3 Cartan-Schouten Connection

The conclusion of the Chapter is done giving the definition of Cartan-Schouten connection on Lie group, which has the property of being bi-invariant. Indeed for Lie groups left or right invariant metric provide a nice

setting as the Lie group becomes a geodesically complete Riemannian manifold, thus also metrically complete. This Riemannian approach is fully consistent with the group operations only if a bi-invariant metric exists. More details are shown in [8].

Definition 1.42. For any vector fields X and Y and any group element $g \in G$ we say that a connection is a *left invariant connection* if satisfies $\nabla_{\mathrm{d}\ell_g X}\mathrm{d}\ell_g Y = \mathrm{d}\ell_g \nabla_X Y$

Definition 1.43. Among the left invariant connections, the *Cartan-Schouten connections* are the ones for which geodesics going through identity are one-parametry subgroups. Bi-invariant connections are both left and right invariant.

Theorem 1.44. *Cartan-Schouten connections are uniquely determined by the property $\alpha(x,x) = 0$ for all $x \in \mathfrak{g}$ where $\alpha : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$.*
Bi-invariant connections are characterizeed by the condition:

$$\alpha([z,x],y) + \alpha(x,[z,y]) = [z,\alpha(x,y)] \quad \forall x,y,z \in \mathfrak{g}$$

The one dimensonal family of connections generated by $\alpha(x,y) = \lambda[x,y]$ satisfy these two conditions. Moreover, there is a unique symmetric Cartan-Schouten bi-invariant connection called the canonical connection of the Lie group (also called mean or $0-$connection) defined by $\alpha(x,y) = \frac{1}{2}[x,y]$ for all $x,y \in \mathfrak{g}$ i.e. $\nabla_{\tilde{X}}\tilde{Y} = \frac{1}{2}[\tilde{X},\tilde{Y}]$ for two left-invariant vector fields.

Chapter 2

Statistics on Riemmannian Manifolds

The purpose of this chapter is to define the statistics which are needed to explain Barycentric Subspaces. In this studies the data lie on a known manifold, and the goal is to study statistics of data restricted to this manifold. Then the statistical computing on manifolds is a domain in which differential geometry meets statistics.
Information about probability theory is taken from Pennec [10].

Firstly the definition of probability space has to be transferred on a Riemannian manifold. For doing that it is needed a definition of measure on the manifold and this is induced by the Riemannian metric, i.e. by the infinitesimal volume element on each tangent space:

Definition 2.1. The volume element in n-dimensional Riemannian manifold with metric $G(x) = [g_{ij}(x)]$ is defined by the following formula:

$$\mathrm{d}\mathcal{M}(x) = \sqrt{|\det G(x)|}\mathrm{d}x$$

Let $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ be a probability space where Ω is the whole space of the events, $\mathcal{B}(\Omega)$ is the σ-algebra of Borel, i.e. the smallest σ-algebra containing all the open subsets of Ω and $\mathbb{P} : \mathcal{B}(\Omega) \to [0, 1]$ is a probability measure, i.e. $\mathbb{P}(\Omega) = 1$ and it has to satisfy Kolmogorov's axioms.

Definition 2.2. A random point in the Riemannian manifold $\mathcal{M}$ is a Borel measurable function $X = X(w)$ from Ω to $\mathcal{M}$.

The induced measure is $\mathbb{P} \circ X^{-1}$. In particular, take

$$X : (\Omega, \mathcal{B}, \mathbb{P}) \to (\mathcal{M}, \mathcal{A})$$

where $\mathcal{A}$ is the $\sigma-$algebra on $\mathcal{M}$.

It is important to define the probability of an event occurring, this depends on its distribution. This can be given by its distribution function, by its probability mass function (if the variables are discrete) or by probability density function (if the variables are continuous). So in case the variables are continuous, denoting the probability of an event χ occurring by $\mathbb{P}(X \in \chi)$:

Definition 2.3. Let $\mathcal{A}$ be the Borel $\sigma-$algebra of $\mathcal{M}$. The random point X has probability density function p_X (real positive and integrable function) if:

$$\forall\, \chi \in \mathcal{A} \quad \mathbb{P}(X \in \chi) \int_{\chi} p(y)\mathrm{d}\mathcal{M}(y) \quad \mathbb{P}(X \in \mathcal{M}) = \int_{\mathcal{M}} p(y)\mathrm{d}\mathcal{M}(y) = 1$$

Since the cut locus has null measure it is possible to integrate on $\mathcal{M}$ in an exponential chart.
If f is an integrable function of the manifold and $f_p(\vec{pq}) = f(\exp_p(\vec{pq}))$ is its image in the exponential chart at p, we have:

$$\int_M f(p)\mathrm{d}M = \int_{\mathcal{D}(p)} f_p(\vec{z})\sqrt{G_{\vec{x}}(\vec{z})}\mathrm{d}\,\vec{z}$$

2.1 Expected value of a function

One of the main concept to define in statistic is the expected value, i.e. the first moment of a distribution. Moments are really important to distinguish one distribution from another.
Let $\phi(X(w))$ be a Borelian real valued function defined on $\mathcal{M}$ and χ a random point of pdf f_x. Then $\phi(\chi)$ is a real random variable and we can compute its expectation:

$$E[\phi(X)] = \int_{\mathcal{M}} \phi(y)p_X(y)\mathrm{d}\mathcal{M}(y)$$

2.2 Variance

The second moment of a distribution is the variance, or the square of the standard deviation, which is very important in statistic because it measures the spread of the data

$$\sigma_X^2(y) = E[\mathrm{dist}(y, X)^2] = \int_{\mathcal{M}} \mathrm{dist}(y, z)^2 p_X(z)\mathrm{d}\mathcal{M}(z)$$

2.3 Fréchet expectation

One of the most interesting definition of expected value for geodesically complete Riemannian manifolds is the Fréchet and Karcher expectation. It is important to notice that the definition can be used in metric space, which are definitly more general than manifolds.

Definition 2.4 (Fréchet expectation)**.** Let X be a random point. If the variance $\sigma_X^2(y)$ is finite for all point $y \in \mathcal{M}$, every point minimizing this variance is called expected or mean point. Thus, the set of mean points is:

$$\mathbb{E}[X] = \arg\min_{y \in \mathcal{M}} (E[\mathrm{dist}(y, X)^2])$$

if there exists a least one mean point $\overline{x}$, it is called variance the minimal value $\sigma_X^2 = \sigma_X^2(\overline{x})$.

In case of a set of discrete measures $X_1, \cdots, X_n$ the *empirical or discrete mean points*:

$$\mathbb{E}[\{X_i\}] = arg\min_{y \in \mathcal{M}} (E[\left\{\mathrm{dist}(y, X_i)^2\right\}]) = arg\min_{y \in \mathcal{M}} \left(\frac{1}{n} \sum_i \mathrm{dist}(y, X_i)^2 \right)$$

So the Fréchet are the set of points minimizing globally the variance, while Karcher proposed to consider the local minima. Thus, the Fréchet mean are the subset of the Karcher ones.

2.3.1 Metric power of Fréchet expectation

The Fréchet (resp. Karcher) mean can be further generalized by taking a power α of the metric define the $\alpha-$variance.

$$\sigma_{X,\alpha}(y) = (E[\mathrm{dist}(y, X)^\alpha])^{\frac{1}{\alpha}} = \left(\int_{\mathcal{M}} \mathrm{dist}(y, z)^\alpha p_X(z) \mathrm{d}\mathcal{M}(z) \right)^{\frac{1}{\alpha}}$$

In case of measures $X_1, \cdots, X_n$ the *empirical or discrete mean point*:

$$\sigma_{X,\alpha}(y) = (E[\mathrm{dist}(y, X)^\alpha])^{\frac{1}{\alpha}} = \left(\frac{1}{n} \sum_i \mathrm{dist}^\alpha(y, X_i) \right)^{\frac{1}{\alpha}}$$

The global minima of the $\alpha-$variance defines the Fréchet median for $\alpha = 1$, the Fréchet mean for $\alpha = 2$ and the barycenter of the support of the distribution for $\alpha = \infty$.

Chapter 3

Barycentric Subspace Analysis

A powerful tool for multivariate statistical analysis is *principal component analysis* (PCA), which is very important to reduce the dimensionality of the data and yields a hierarchy of major directions explaining the main sources of data variation. The problem arise when multivariate data lies in a non-Euclidean space, such as Riemannian structure. Then it is needed to generalize this tool.
In this chapter it is introduced the definition of Barycentric Subspace, which was first introduced by X. Pennec (2015) and it is a generalization of the concept of affine subspace in Euclidean space. Depending on the generalization of the mean that is used on the manifold: Fréchet and Karcher mean or exponential barycenter, it is obtained the Fréchet/Karcher subspaces or the affine span. These three definitions are related. Barycentric Subspace Analysis (BSA) is then generalization of PCA done building a forward or backward analysis of nested subspaces. All the information are taken from [7].

3.1 Barycentric Subspace

First of all, the setting of the work is a Riemannian manifold geodesically complete. Then, since in a Riemannian manifold the Riemannian log and distance functions are not smooth in the cut locus, it is necessary to give two definitions for working on the manifold:

Definition 3.1. Let $\{x_0, \cdots, x_k\} \subset \mathcal{M}$ be a set of $k+1 \leq n$ reference points in the n dimensional Riemannian manifold $\mathcal{M}$ and $\mathcal{C}(x_0, \cdots, x_k) = \bigcup_{i=0}^{k} \mathcal{C}(x_i)$ be the union of the cut loci of these points. It is called the smooth manifold $\mathcal{M}$ and the $k+1$ reference points a $(k+1)$-*pointed manifold.* The submanifold

$\mathcal{M}^*(x_0,\cdots,x_k) = \mathcal{M} \setminus \mathcal{C}(x_0,\cdots,x_k)$ of the non-cut points of the $k+1$ reference points it is called $(k+1)$-*punctured manifold.*

Thank to that definition, the Riemannian log is well defined for all the points of the $(k+1)$-punctured manifold. Moreover, since the cut locus has null measure $\mathcal{M}^*(x_0,\cdots,x_k)$ is open and dense in $\mathcal{M}$, then it is a submanifold of $\mathcal{M}$,but not necessary connected.

Definition 3.2. Let $(\lambda_0,\cdots,\lambda_k) \in \mathbb{R}^{k+1}$ such that $\sum_i \lambda_i \neq 0$. The quotient $\underline{\lambda_i} = \frac{\lambda_i}{\sum_{j=0}^k \lambda_j}$ is called the normalized weights. The weighted $p-$order moment of a $(k+1)$-pointed Riemannian manifold is the p-contravariant tensor:

$$\mathfrak{M}_p(x,\lambda) = \sum_i \lambda_i \underbrace{\overrightarrow{xx_i} \otimes \overrightarrow{xx_i} \cdots \otimes \overrightarrow{xx_i}}_{p\,\text{times}}$$

and the normalized weighted p-order moment is:

$$\underline{\mathfrak{M}}_p(x,\lambda) = \mathfrak{M}_p(x,\underline{\lambda}) = \frac{\mathfrak{M}_p(x,\lambda)}{\mathfrak{M}_0(\underline{\lambda})}$$

For a fixed weight λ, the first order $\mathfrak{M}_1(x,\lambda) = \sum_i \lambda_i \overrightarrow{xx_i}$ is a smooth vector field on the manifold $\mathcal{M}^*(x_0,\cdots,x_k)$.

Remark 3.3. Let $(\mathcal{M},g)$ be a Riemannian manifold and x_0,x_1,x_2 three reference points, the first moment is:

$$\mathfrak{M}_1(x,\lambda) = \lambda_0 \log_x(x_0) + \lambda_1 \log_x(x_1) + \lambda_2 \log_x(x_2)$$

So the reference points span a 3-dimensional subspace of $T_x\mathcal{M}$ for arbitrary weight λ.

From Euclidean affine subspaces

In the Euclidean space the barycenter p of $x_1,\cdots,x_k$ points is defined as:

$$p = \sum_{i=1}^k \lambda_i x_i$$

where $\sum_{i=1}^k \lambda_i = 1$. In particular p is independent on the choice of the origin x_0:

$$\begin{aligned} p &= \sum_{i=1}^k \lambda_i x_i \\ p - x_0 &= \sum_{i=1}^k \lambda_i x_i - \sum_i \lambda_i x_0 \\ p - x_0 &= \sum_{i=1}^k \lambda_i (x_i - x_0) \end{aligned}$$

This definition can be reinterpreted, we call $p = x_0$. Then x_0 is the barycenter if

$$\sum_{i=1}^{k} \lambda_i (x_i - x_0) = 0$$

with $\sum_{i=1}^{k} \lambda_i = 1$.
Then the vector $\lambda = (\lambda_1, \cdots, \lambda_k)$ is seen as a vector orthogonal at the space where the points lie. As λ is giving a direction and this is the reason why it stay in the projective space. In other words the barycenter lie in the subspace generated by the given points.

Example 3.4. Thinking in $\mathbb{R}^2$ taking two different points $x_1 = (2, 0)$ and $x_2 = (1, 3)$ looking for the barycentric coordinate of the point $x = (x_0, y_0)$ it means:

$$\lambda_1 \begin{pmatrix} 2 - x_0 \\ 0 - y_0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 - x_0 \\ 3 - y_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

If the point is the origin $x_0 = (0, 0)$:

$$\lambda_1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The solution λ of the system for Cramer are $\lambda_1 = \lambda_2 = 0$ but they are not an admissible choices.

Example 3.5. Always in $\mathbb{R}^2$ taking two different points $x_1 = (-1, 1)$ and $x_2 = (1, 1)$ looking for the barycentric coordinate of the point $x = (0, 1)$ means:

$$\lambda_1 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The solution λ of the system for Cramer are $\lambda_1 = \lambda_2$. It is remarkable that the point is not necessary the barycenter, since there are barycentric coordinates even if the point lie in the line but not in the barycentric point.

This is the reason to define the barycentric coordinates on the manifold:

Definition 3.6 (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 it has been removed the codimension 1 subspace $\mathbf{1}^\perp$ orthogonal to the point $\mathbf{1} = (1 : 1 : \cdots : 1)$:

$$\mathcal{P}_k^* = \left\{ \lambda = (\lambda_0 : \lambda_1 : \cdots : \lambda_k) \in \mathbb{R}P^n \quad \text{s.t.} \, \mathbf{1}^T \lambda \neq 0 \right\}$$

Projective points are represented by lines through the origin. Standard representations of this space are given by the intersection of the lines with the "upper" unit sphere S^k of $\mathbb{R}^{k+1}$ with north pole $\mathbf{1}/\sqrt{k+1}$ or by the affine k–plane of $\mathbb{R}^{k+1}$ passing through the point $\mathbf{1}/(k+1)$ and orthogonal to this vector. This last representation amounts to use the normalized weights $\underline{\lambda}_i = \lambda_i/(\sum_{j=0}^{k}\lambda_j)$, for which the vertices of the simplex have homogeneous coordinates $(1,0,\cdots,0)\cdots(0,0,\cdots,1)$.

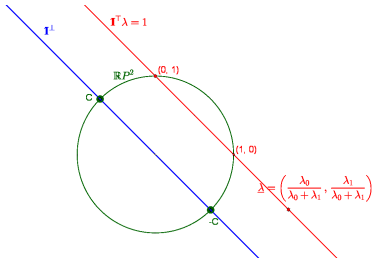


Figure 3.1: Projective weights for $k = 1$. Image taken from [7]

To prevent weights to sum up to zero, it was necessary to remove the codimension 1 subspace $\mathbf{1}^\perp$ orthogonal to the projective point $\mathbf{1} = (1:1:\cdots,1)$. This excluded subspace corresponds to the equator of the pole $\mathbf{1}/\sqrt{k+1}$ for the sphere representation (points C and $-C$ identified), and to the projective completion (points at infinity) of the affine k–plane of normalized weights.

Definition 3.7 (Barycentric coordinates in a $(k+1)$pointed manifold)**.** A point $x \in \mathcal{M}^*(x_0,\cdots,x_k)$ has barycentric coordinates $\lambda \in \mathcal{P}_k^*$ if:

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

As said previously, this definition is well posed for the punctured manifold $\mathcal{M}^*(x_0,\cdots,x_k)$.

As a consequence, now it is given a definition of subspaces on a manifold which restores the full symmetry between all the parameters without privileging one point:

Definition 3.8 (Exponential Barycentric Subspace(EBS))**.** The *Exponential Barycentric Subspace(EBS)* of the points $(x_0,\cdots,x_k) \in \mathcal{M}^{k+1}$ is the locus of weighted exponential barycenters of the reference points in $\mathcal{M}^*(x_0,\cdots,x_k)$:

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

In the following chapter it will be shown that in the sphere this definition means to find the great circle passing by the reference points. It was seen in the Euclidean space that there were not a privileged points, and this is the property that we are looking for on a manifold, for instance in the sphere all the points belonging to the subspace are the one in the great circle passing by the reference points, and then there are infinite points that can be chosen as central points.

As a consequence, the space of admissible barycentric weights is

$$\Lambda(x) = \{\lambda \in \mathcal{P}_k^* \,|\mathfrak{M}_1(x,\lambda) = 0\}$$

Proposition 3.9. *$x \in EBS(x_0,\cdots,x_k)$ if and only if $\Lambda(x) \neq \varnothing$*

The discontinuity of the Riemannian log on the cut locus of the reference points may hide continuity or discontinuities of the exponential barycentric subspace. Then for ensure the completeness and reconnect different components, it is considered the closure of this set:

Definition 3.10 (Affine span)**.** The *affine span* is the closure of the Exponential Barycentric Subspace in M:

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

Since $\mathcal{M}$ is geodesically complete, the metric completion of the EBS is guaranteed.

The completeness of the affine span is fundamental because it allows to always find a closest point of the data on the subspace which is really important in the practice.

Example 3.11. In $\mathbb{R}^2$ taken two reference points $x_1 = (1,1)$ and $x_2 = (0,1)$, the $EBS(x_1,x_2) = \{x \in \mathbb{R}^2 | \exists \lambda \in \mathcal{P}^* \; \mathfrak{M}_1(x,\lambda) = 0\}$.

$$\lambda_1 \begin{pmatrix} x-1 \\ y-1 \end{pmatrix} + \lambda_2 \begin{pmatrix} x-0 \\ y-1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Then:

$$\begin{cases} \lambda_1(x-1) + \lambda_2 x = 0 \\ \lambda_2(y-1) + \lambda_2(y-1) = 0 \end{cases}$$

Solving the system for x and y:

$$\begin{cases} \lambda_1 x - \lambda_1 + \lambda_2 x = 0 \\ \lambda_2 y - \lambda_2 + \lambda_2 y - \lambda_2 = 0 \end{cases}$$

$$\begin{cases} x = \frac{\lambda_1}{\lambda_1+\lambda_2} \\ y = 1 \end{cases}$$

The solution is

$$EBS(x_1, x_2) = \left\{ (x, y) \in \mathbb{R}^2 | (\frac{\lambda_1}{\lambda_1 + \lambda_2}, 1) \right\}$$

Since, the setting is $\mathbb{R}^2$ the solution is the straight line passing through the two reference points, i.e. classical affine subspace in the Euclidean Space.

Let $Z(x) = [\vec{xx}_0, \cdots, \vec{xx}_k]$ be the smooth field of $n \times (k+1)$ matrices of vectors pointing from any point $x \in \mathcal{M}^*(x_0, \cdots, x_k)$ to the reference points. The constraint $\sum_i \lambda_i \ \vec{xx}_i = 0$ can be rewritten in matrix form $\mathfrak{M}_1(x, \lambda) = Z(x)\lambda = 0$ where λ is the $k+1$ vector of homogeneous coordinates λ_i.
A direct consequence of this statement is the following theorem:

Theorem 3.12. *Let $Z(x) = U(x)S(x)V(x)^T$ be a singular decomposition of the $n \times (k+1)$ matrix fields $Z(x) = [\vec{xx}_0, \cdots, \vec{xx}_k]$ on $M^*(x_0, \cdots, x_k)$ with singular values $\{s_i(x)\}_{0 \le i \le k}$ sorted in decreasing order. $EBS(x_0, \cdots, x_k)$ is the zero levelset of the smallest singular value $s_k(x)$ and the dual subspace of valid barycentric weights is spanned by the right singular vectors corresponding to the l vanishing singular values: $\Lambda(x) = Span(v_{k-l}, \cdots, v_k)$ (it is void if $l = 0$)*

Proof. Following the paper [7]. 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: $\Lambda(x) = Ker(Z(x))$. If there is only one zero singular value ($s_k = 0$ and $s_{k-1} > 0$), then λ is proportional to v_k. If l singular values vanish, then there is a higher dimensional linear subspace of solution for λ. □

Example 3.13. Let consider in $\mathbb{R}^5$ the two points: $x_0 = (1,0,0,0,0)$ and $x_1 = (0,1,0,0,0)$, then the matrix of them is $Z(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$. The singular values of $Z(x)$ are the square root of the eigenvalues of $Z(x)^T Z(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Then the solutions are $s_0 = s_1 = 1$ and the other are zero. Then $s = \text{diag}(1,1,0,0,0)$. The theorem tells us that the subspace of valid barycentric weights is $\Lambda(x) = \text{Span}(e_3, e_4, e_5)$ and the EBS is the orthogonal $e_3 = e_4 = e_5$.

Theorem 3.14. *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 $M^*(x_0, \cdots, x_k)$ with components $\Omega_{ij}(x) =< \vec{xx}_i, \vec{xx}_j >$ and $\Sigma(x) = \mathfrak{M}_2(x, 1) = \sum_{i=0}^{k} \vec{xx}_i \vec{xx}_i^T = Z(x)Z(x)^T$ be the (scaled) $n \times n$ covariance matrix field of the reference points. $EBS(x_0, \cdots, x_k)$ 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. Following the proof from [7].
The constraint $\mathfrak{M}_1(x, \lambda) = 0$ is satisfied if and only if:

$$\|\mathfrak{M}_1(x, \lambda)\|_x^2 = \| \sum_i \lambda_i \overrightarrow{xx_i} \|_x^= \lambda^T . \Omega(x) . \lambda = 0$$

As the function is homogeneous in λ, it can be restricted to unit vectors. Adding this constrains with a Lagrangian multiplier to the cost function, it is ended-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 there is $\|\mathfrak{M}_1(x, \mu_{k+1}(x))\|_2^2 = \sigma_{k+1}(x)$, 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 $\Omega(x)$:

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

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 non singular but non orthogonal matrix) obviously changes the singular values in general. There is however a special case where some singular values are equal: this is 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 (it was intentionally dropped the usual normalization factor $1/k+1$ to simplify the notations), let us notice first that the definition does not assumes that the coordinate system is orthonormal. Thus, the eigenvalues of the covariance matrix are depending 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 last decomposition is sometimes called the proper orthogonal decomposition (POD). Now, 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, barycentric subspace may also be characterized as the zero level-set of the $k+1$ eigenvalue (sorted in decreasing order) of Σ (or of the joint eigenvalue of $\Sigma(x)$ and $G(x)$)), and this characterization is once again independent of the basis chosen. □

3.1.1 Fréchet and Karcher Barycentric subspaces in metric spaces

Using the Fréchet or Karcher definition of the mean it can be obtained another kind of barycentric subspaces, defined as the locus of weighted Fréchet (or Karcher) means.

Definition 3.15 (Fréchet and Karcher barycentric subspaces of $k+1$ points)**.** Let $(\mathcal{M}, \mathrm{dist})$ a metric space of dimensione n and $(x_0, \cdots, x_k) \in \mathcal{M}^{k+1}$ be $k+1 \leq n$ distinct reference points. The (normalized) weighted variance at point x with weight $\lambda \in \mathcal{P}_k^*$ is $\sigma^2(x, \lambda) = \frac{1}{2}\sum_{i=0}^k \underline{\lambda}_i \mathrm{dist}^2(x, x_i) = \frac{1}{2}\sum_{i=0}^k \mathrm{dist}^2(x, x_i) / \sum_{j=0}^k \lambda_j$. The *Fréchet barycentric subspace* of these points is the locus of weighted Fréchet means of these points, i.e. the set of the absolute minima of the weighted variance:

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

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

It is really important to notice that these definitions work on metric space which are more general than Riemannian space.

Link between the difference barycentric subspaces

Firstly it is clear that the locus of local minima of the variance (i.e. Karcher mean) is a superset of global minima (Fréchet mean).
Moreover on the punctured manifold $\mathcal{M}(x_0, \cdots, x_k)$ the squared distance $\mathrm{d}^2_{x_i}(x) = \mathrm{dist}^2(x, x_i)$ is smooth. Then it can be computed its gradient $\nabla \mathrm{d}^2_{x_i}(x) = -2\log_x(x_i)$.
Hence, the relationship with the EBS appears, indeed the EBS equation is the sum of the weighted Riemannian log: $\sum_i \underline{\lambda}_i \log_x(x_i) = 0$ so it defines the critical points of the weighted variance:

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

Stability of affine subspace with different metric power

In the previous chapter in section 2.3.1 the Fréchet and Karcher mean definition was generalized by taking the α-power of the variance. In the same way it can be taken the α-variance of Fréchet and Karcher to give a general definition of barycentric subspaces.
It turns out that these α–subspaces are necessary included in the affine span. Indeed:

$$\nabla_x \sigma^\alpha(x, \alpha) = -\sum_{i=0}^{k} \underline{\lambda}_i \mathrm{dist}^{\alpha-2}(x, x_i) \log_x(x_i)$$

The critical points of the α-variance are simply elements of the EBS and changing the power of the metric just amounts to a reparametrization of the barycentric weights. Then the stability of affine span with respect to the power shows that the affine span is really a central notion.

3.2 Barycentric Subspace Analysis

One of the main property satisfy by the Euclidean PCA is to create nested linear spaces that best approximate the data at each level. This is a really interesting property that leads to use the barycentric subspaces as a generalization of PCA thanks to their property of being easily nested obtaining a family of embedded submanifolds which generalizes flags of vector spaces.
A strict ordering of n+1 independent points $x_0 \prec x_1 \cdots \prec x_n$ on an n–dimensional manifold $\mathcal{M}$ defines the filtration of subspaces for Barycentric Subspaces, for instance $\mathrm{EBS}(x_0) = \{x_0\} \subset \cdots \mathrm{EBS}(x_0, x_1, \cdots, x_k) \cdots \subset \mathrm{EBS}(x_0, \cdots, x_n)$.
It was noticed in section 3.1.1 that the most appealing definition was the affine span. Indeed, if the manifold is connected the EBS of $n+1$ distinct points covers the full manifold $\mathcal{M}^*(x_0, \cdots, x_k)$, and as a consequence the affine span covers the whole original manifold $Aff(x_0, \cdots, x_n) = \mathcal{M}$. Clearly the Fréchet or Karcher barycentric subspaces generate only a submanifold that does not cover the whole manifold in general.
Then it is given the proper definition of flags:

Definition 3.16 (Flags of affine spans in manifolds)**.** Let $x_0 \prec x_1 \cdots \prec x_k$ be $k+1 \leq n$ distinct and (non-strictly) ordered points of $\mathcal{M}$. By non-striclty, it means that 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 $\mathrm{FL}_i(x_0 \prec x_1 \cdots \prec x_k) = \mathrm{Aff}(x_0, \cdots, x_i)$ for $0 \leq i \leq k$ the flag of affine spans $\mathrm{FL}(x_0 \prec x_1 \cdots \prec x_k)$.
For non strictly ordered sets of points $x_0 \prec x - 1 \cdots \prec x_k$, subspaces in the

sequence are only generated at strict ordering signs or at the end, so that all exchangeable points are always considered together. A flag of exchangeable points $\mathrm{FL}(x_0 \sim x_1 \cdots \sim x_k) = \mathrm{Aff}(x_0, \cdots, x_k)$.

In conclusion it will be explained the three methods to build a sequence of barycentric subspaces.

3.2.1 Forward barycentric subspaces analysis

In general a forward method means that it starts with a 0-dimensional space and it grows of one dimension each step until it covers the whole space. Indeed the *forward barycentric subspace analysis* starts exactly computing the point which is the optimal barycentric subspace generated by only one point $\mathrm{Aff}(x_0) = \{x_0\}$ minimizing the unexplained variance, i.e. the Karcher mean.
Adding a second point the goal is to find the first-dimensional subspace generated by the two points which is the geodesic passing through them: the first point is the Karcher mean found before, while the second is chosen as the best point for whom the geodesic best fit the data. The third step implies to add another point that with the other two best explain the data with this construction the Fréchet mean is always belonging to the subspace.
The stopping criterion for this subspaces can be chosen or fixing a maximum number of subspaces or when the variance of the residues reaches a threshold. One of the problem of a forward approach is that it is a greedy algorithm, since the previous points are fixed, it is not ensured that the ones chosen to build the $k-$dimensional affine span are the best to fit the data also for the $k+1$-dimensional affine span. In other words the affine span of dimension k defined by the first $k+1$ points is not in general the optimal one minimizing the unexplained variance.

We compute the computational cost of the $k-$th step to built FBS. To simplify the calculation we restrict to the case where the choice of the reference points is limited to the actual data points so we are in a sample-limited optimization. Then we consider k reference points among n data points, then compute the projection of $n-k$ points and take the distance to project, we call π the cost of minimizing the distance. So the cost is given by:

$$\mathcal{C}_{K^{th}}^{FBS} = (n+1-k)(n-k)\pi \tag{3.1}$$

The first term specifies that we are looking for the $k-$point to built the subspace but the previous are fixed. The other two terms show the cost of computing the unexplained variance.

However the total cost of the $k-$th subspace is computed summing up all the previous step:

$$\mathcal{C}_k^{FBS} = \sum_{j=1}^{k}(n+1-j)(n-j)\pi \approx \mathcal{O}(kn^2)$$

3.2.2 Backward barycentric subspaces analysis or Pure Barycentric Subspace

On the other hand the backward analysis usually starts with the complete manifold and at each step removed a point. Then it is needed an optimization method to chose which point has to be removed.
In practice the optimization is done on the $k+1$ points to find the k dimensional affine span, and then reorder the points using a backward sweep to find inductively the one that at least increase the unexplained variance. This method is called 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 do.
For instance to compute $k-$dimensional PBS all the combination of $k+1$ points of the data is computed and then is add an order on the points. In this way it seems that k-PBS fit better than k-FBS, because in FBS, it is only needed to add the k-th point while the other $(k+1)$points are fixed. However in PBS, after have found the points it is put an order on them. This means that going in a backward analysis it is needed to follow this order,and this is not always fitting well the data.

The complexity of PBS using the sample-limited case and the same notation of 3.1:

$$\mathcal{C}_k^{PBS} = \binom{n}{k}(n-k)\pi \approx \mathcal{O}(n^{k+1})$$

the first term is related to all the possible combination of $k-$points between the n data points we have.

3.2.3 A criterion for hierarchies of subspaces

The first two analysis presented used the unexplained variance as a criterion to build the subspaces, but to obtain consistency across dimensions, it is better to find another criterion which depends on the whole flag of subspaces. It can be defined the *Accumulated-Unexplained-Variance* (AUV).
Given a strictly ordered flag of affine subspaces $Fl(x_0 \prec x_1 \cdots \prec x_k)$ the *AUV*

criterion:

$$AUV(Fl(x_0 \prec x_1 \cdots \prec x_k)) = \sum_{i=0}^{k} \sigma^2(Fl(x_0 \prec x_1 \cdots \prec x_k)) \tag{3.2}$$

With this global criterion the point x_i influences all the subspaces of the flag that are larger than $Fl_i(x_0 \prec x_1 \cdots \prec x_k)$ but not the smaller subspaces. This results lead into a particularly appealing generalization of PCA on manifolds called *Barycentric Subspaces Analysis* (BSA).

3.2.4 Barycentric Subspace Analysis

The Barycentric Subspace Analysis use the AUV criterion described to compute the $k-$th susbspace a sthe one which minimizes AUV. For instance, finding 1-BSA means to test all the possible couple of points (x_0, x_1) of the Data, but in this case the order is very important, because testing the couple (x_0, x_1) means minimize $AUV(x_0 \prec x_1) = \text{dist}^2(x_0, Data) + \text{dist}^2([x_0, x_1], Data)$ which is really different from the couple (x_1, x_0).
The complexity of BSA using the sample-limited case and the same notation of 3.1:

$$\mathcal{C}_k^{BSA} = \binom{n}{k} \sum_{j=1}^{k} (n-j)\pi \approx \mathcal{O}(kn^{k+1})$$

The first term want to try all the possible combination of $k-$points between all the n data points, while the last part of the formula show the cost of the criterion depending on the whole flag.
In conclusion, three different way to build the barycentric subspaces are given. The Forward and the Pure use the unexplained variance as the criterion while the Barycentric one use a criterion based on all the previous subspaces. This suggest that the last one is going to better fit the data than the other two, but the computational cost will be higher.

Chapter 4

Barycentric Subspace Analysis applied on the Sphere

In the first part of the Chapter is presented the probability density functions space [1] [2] it leads to work on the Sphere. Indeed the Sphere is a really interesting manifold, it is nonlinear, it is simple, it has constant curvature and it is well known. For this reasons the second part of the Chapter presents a deep study on the Sphere and the three different type of Barycentric Subspaces are tested, the full code is available in the Appendix. We start by an example coming from information geometry to show the importance of the sphere.

4.1 The space of probability density functions

Information geometry is a branch of mathematics that applies the techniques of differential geometry to the field of probability theory. This is done by taking probability distributions for a statistical model as the points of a Riemannian manifold, forming the statistical manifold.
The probability density functions (pdf) are used for example in modeling frequencies of pixel values in images. An important step in classifying observations using these functions is to compute distances between any two arbitrary functions.
Let $\mathcal{P}$ be the space of probability density functions p which are defined, for simplicity, on the interval $[0,1]$:

$$\mathcal{P} = \left\{ p : [0,1] \to \mathbb{R} \,|\, \forall s \quad p(s) \geq 0 \quad \text{and} \int_0^1 p(s)\mathrm{d}s = 1 \right\}$$

On this space, which is not a vector space, the natural metric is the Fisher-Rao metric, the L^2 norm, which is defined over the tangent space $v_1, v_2 \in$

$T_p(\mathcal{P})$ for each point $p \in \mathcal{P}$ as:

$$< v_1, v_2 >= \int_0^1 v_1(s)v_2(s)\frac{1}{p(s)}\mathrm{d}s$$

The main problem is that working with this representation leads to some difficulties. It turns out that if the representation choice for describing this space is the square root functions: $\psi = \sqrt{p}$. Then the space to consider is:

$$\Psi = \left\{\psi : [0,1] \to \mathbb{R} \,|\, \psi \geq 0 \quad \text{and} \quad \int_0^1 \psi^2(s)\mathrm{d}s = 1\right\}$$

For any two tangent vectors $v_1, v_2 \in T\psi(\Psi)$ the Fisher-Rao metric:

$$< v_1, v_2 >= \int_0^1 v_1(s)v_2(s)\mathrm{d}s$$

This space can be viewed as the non-negative orthant of the unit Sphere in a Hilbert space, and this is another interesting reason to study in more details the Sphere.
Moreover in [1] it is shown that on a closed manifold of dimension greater than one, every smooth weak Riemannian metric on the space of smooth positive probability densities, invariant under the action of diffeomorphism group, is a multiple of the Fisher-Rao metric.
Indeed it has been proven that this metric is the natural metric for this space and the proof was first done for the finite dimensional submanifolds and after for infinite dimensional manifold of all positive probability densities.
To conclude the section the following theorem is stated:

Theorem 4.1. *The Fisher-Rao metric is invariant to reparametrizations.*

Proof. From [2].
Let $v_1, v_2 \in T\psi(\Psi)$ for some $\psi \in \Psi$ and $\phi \in \Phi$ be a a diffeomorphic function. The re-parametrization action takes ψ to $\psi(\phi)\sqrt{\dot{\phi}}$ and v_i to $\tilde{v_i} \equiv v_i(\phi)\sqrt{\dot{\phi}}$. The inner product after re-parametrization is given by:

$$\int_0^1 \tilde{v_1}(s)\tilde{v_2}(s)\mathrm{d}s = \int_0^1 v_1(\phi(s))\sqrt{\dot{\phi(s)}}v_2(\phi(s))\sqrt{\dot{\phi(s)}}\mathrm{d}s = \int_0^1 v_1(t)v_2(t)\mathrm{d}t$$

where $t = \phi(s)$. which is the same before re-parametrizaion and, hence, invariant. □

4.2 the Sphere

The n–dimensional Sphere embedded in $\mathbb{R}^{n+1}$ is defined as

$$S^n = \left\{ (x_1, \cdots, x_{n+1}) \in \mathbb{R}^{n+1} \text{ s.t. } \sum_{i=1}^{n+1} x_i^2 = 1 \right\}$$

It is a manifolds for the Theorem 1.3. It is a really important example of non-linear space, since for every $p, q \in S^n$ the sum $p + q$ is not in S^n.
The tangent space of the Sphere is defined as $T_x S^n = \{v \in \mathbb{R}^{n+1},\ v^T x = 0\}$ and the scalar product defined on it which defines the Riemannian metric is inherited from the Euclidean metric, as it was shown in Chapter 1.
The Riemannian distance is then $d(x, y) = \arccos(x^T, y) = \theta$ with $\theta \in [0, \pi]$ and the geodesics are the great circle passing through two points.
One of the most important things to implement that allows to move from the manifold S^n and its tangent space are the Spherical Exponential and Logarithmic maps. Then the two maps are given by the following two formula:

$$\exp_x(v) = \cos(||v||)x + \operatorname{sinc}(||v||)v$$

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

where $f :]-\pi, \pi[\to \mathbb{R} \in \mathcal{C}^\infty$ and it is defined as $f(\theta) = \frac{1}{\operatorname{sinc}\theta} = \frac{\theta}{\sin\theta}$.
To implement the Spherical Exponential map, it is needed to check if the starting vector v belongs to the tangent space, so the projection in the tangent space is computed: $w \in T_x S^n$: $w = (v - < x, v > x)$. Then:

$$\exp_x : T_x S^n \to S^n$$
$$w \to \cos(\sqrt{w^T w})x + \frac{\sin(\sqrt{w^T w})}{\sqrt{w^T w}} w$$

if the norm of w is too near 0, the Taylor expansion is used:

$$w \to \left(1 - \frac{1}{2}w^T w + \frac{1}{24}\left(\sqrt{w^T w}\right)^4 - \frac{1}{720}\left(\sqrt{w^T w}\right)^6 + \frac{1}{40320}\left(\sqrt{w^T w}\right)^8\right)x +$$
$$\left(1 - \frac{1}{6}\left(\sqrt{w^T w}\right)^2 + \frac{1}{120}\left(\sqrt{w^T w}\right)^4 - \frac{1}{5040}\left(\sqrt{w^T w}\right)^6 + \frac{1}{362880}\left(\sqrt{w^T w}\right)^8\right) w$$

The spherical Log:

$$\log_x : S^n \to T_x S^n$$
$$y \to \frac{\theta}{\sin\theta} y - \frac{\theta\cos\theta}{\sin\theta} x$$

if the point is near 0, the Taylor expansion is computed:

$$y \to \left(1 + \frac{1}{6}\theta^2 + \frac{7}{360}\theta^4 + \frac{31}{15120}\theta^6 + \frac{127}{604800}\theta^8\right) y$$
$$\left(1 - \frac{1}{3}\theta^2 - \frac{1}{45}\theta^4 - \frac{2}{945}\theta^6 - \frac{1}{4725}\theta^8\right) x$$

The geometrical setting is then completed and the points of the Sphere represent the data which are the subject of interest of the statistical analysis. In Chapter 3 one of the central definition was the first moment. Selected $k+1$ reference points on the Sphere $(x_0, \cdots, x_k) \in S^n$, the matrix of the reference points is defined as $X = [x_0, \cdots, x_k]$. The cut locus of x_i is its antipodal point $-x_i$ so that the $(k+1)$–punctured manifold is $\mathcal{M}^*(x_0, \cdots, x_k) = S^n \backslash -X$. Denoting $F(X, x) = \text{diag}(f(\arccos(x_i x^t x)))$ the first weighted moment is:

$$\mathfrak{M}_1(x, \lambda) = \sum_i \lambda_i \overrightarrow{xx_i} = (Id - xx^T) X F(X, x) \lambda$$

Let show this formula taken only two reference points $X = [x_1, x_2]$:

$$\begin{aligned}\mathfrak{M}_1(x, \lambda) &= \lambda_1 \log_x(x_1) + \lambda_2 \log_x(x_2) = \\ &= \lambda_1 f(\theta_1)(x_1 - \cos\theta_1 x) + \lambda_2 f(\theta_2)(x_2 - \cos\theta_2 x) \\ &= \lambda_1 f(\theta_1)(x_1 - x^T x_1 x) + \lambda_2 f(\theta_2)(x_2 - x^T x_2 x)\end{aligned}$$

Writing $x^T x_1 = \alpha$ and $x^T x_2 = \beta$ since both $\alpha, \beta \in \mathbb{R}$, it becomes:

$$\begin{aligned}&= \lambda_1 f(\theta_1)(x_1 - \alpha x) + \lambda_2 f(\theta_2)(x_2 - \beta x) \\ &= \begin{pmatrix} x_1 - \alpha x & x_2 - \beta x \end{pmatrix} \begin{pmatrix} f(\theta_1) & 0 \\ 0 & f(\theta_2) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \\ &= (Id - xx^T) \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} f(\theta_1) & 0 \\ 0 & f(\theta_2) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \\ &= (Id - xx^T) X F(X, x) \lambda\end{aligned}$$

Then looking for the $EBS(x_0, \cdots, x_k)$ means to compute where the first moment is zero :

$$(Id - xx^T) X F(X, x) \lambda = 0$$

Since the matrix $F(X, x)$, acting on the homogeneus projective weights, is non-stationary and non-linear in both X and x, it can be simplified by changing the coordinate system with the renormalize weights $\tilde{\lambda} = F(X, x)\lambda$:

$$\begin{aligned}&(Id - xx^T) X \tilde{\lambda} = 0 \\ \Rightarrow &X\tilde{\lambda} - xx^T X \tilde{\lambda} = 0 \\ \Rightarrow &X\tilde{\lambda} = xx^T X \tilde{\lambda}\end{aligned} \tag{4.1}$$

It is remarkable that the left hand side $\alpha = x^T X\tilde{\lambda}$ is a scalar multiple of x, while the right hand side $X\tilde{\lambda}$ is a vector. So the solution can be written requiring that $\alpha \neq 0$:

$$\begin{cases} \alpha = x^T X\tilde{\lambda} \\ x = \frac{X\tilde{\lambda}}{\alpha} \end{cases} \tag{4.2}$$

Since x should live on the $n-$dimensional sphere, it is required also that $\|x\| = 1$. In conclusion the EBS spherical span is:

$$EBS(X) = \text{span}\{x_0, \cdots, x_k\} \cap S^n \backslash X$$

To obtain the affine span, the closure of EBS is taken, which adds the cut locus of the reference points. As a consequence:

$$Aff(X) = \text{span}\{x_0, \cdots, x_k\} \cap S^n$$

Proposition 4.2. *A point x stays in the spherical affine subspace of the matrix $X = [x_0, \cdots, x_k]$ of the reference points if and only if there exists $\overline{\lambda}$ such that $x = X\overline{\lambda}$.*

Theorem 4.3. *The affine span $Aff(X)$ of $k+1 \leq n$ distinct reference unit points $X = [x_0, \cdots, x_k]$ on the $n-$dimensional sphere S^n provided with the canonical metric is the largest subsphere of dimension $Rank(X) - 1$ that contains the reference points.*

4.2.1 Projection onto the affine span

The projection of the Data points onto the affine span is a really crucial computation for the implemenation part.

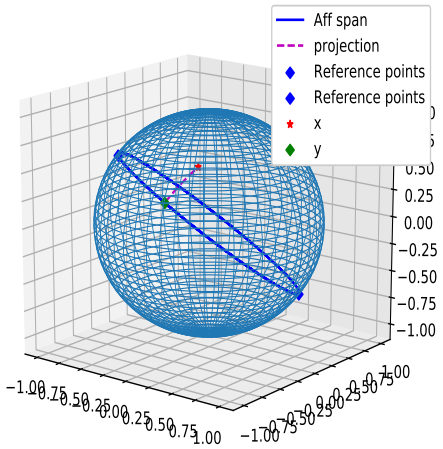


Figure 4.1: The projection y of the point x onto the Affine span generate by two reference points

The goal is to find the projection of the points $x \in S^n$ onto the affine span of the reference points, as done before the matrix of reference points is $X = [x_0, \cdots, x_k]$. The solution is then a point $y \in Aff(X)$ that is the closest point to x which belongs to the subspace.

A Lagrange multiplier is then used to find the point y (Theorem 3.14) which minimizes the distance between x and the subspace under the constrain that $y \in S^n$.

$$\Lambda(y, \alpha) = \mathrm{d}(x, y)^2 + \alpha(||y||^2 - 1)$$

The previous part had proven that a point belongs to the affine span of the reference points only if it exists a λ such that the point can be written as $y = X\lambda$ (proposition 4.2). Moreover asking that the point belongs to the sphere. It can be written the following system:

$y \in \mathrm{Aff}_{span} \Leftrightarrow$

$$\begin{cases} \exists\, \lambda \quad \text{s.t.} \quad y = X\lambda \\ ||y||^2 = 1 \end{cases}$$

Then the Lagrangian becomes:

$$\Lambda(\lambda, \alpha) = \arccos^2(x^T X\lambda) + \alpha(\lambda^T X^T X\lambda - 1)$$

Remark 4.4. It is important to notice before deriving that the matrix X^TX may be rank deficient, if this happen the pseudo inverse of Moore-Penrose can be used.

Deriving for α:

$$\frac{\partial \Lambda}{\partial \alpha} = 0 \Rightarrow \lambda^T X^T X \lambda = 1$$

Deriving for λ :

$$\begin{aligned}
&\frac{\partial \Lambda}{\partial \lambda} = 0 \\
&-\frac{2\arccos(x^T X\lambda)}{\sqrt{1-(x^T X\lambda)^2}} x^T X + 2\alpha\lambda^T X^T X = 0 \\
&-\frac{2\theta}{\sin(\arccos(x^T y))} x^T X + 2\alpha\lambda^T X^T X = 0 \\
&-\frac{2\theta}{\sin\theta} x^T X + 2\alpha\lambda^T X^T X = 0 \\
&\left(\alpha\lambda^T X^T X\right)^T = \left(\frac{\theta}{\sin(\theta)} x^T X\right)^T \\
&\alpha X^T X\lambda = \frac{\theta}{\sin(\theta)} X^T x \\
&\lambda = \frac{\theta}{\sin(\theta)} \frac{1}{\alpha} (X^T X)^{-1} X^T x
\end{aligned}$$

Choosing $\alpha = \frac{\theta}{\sin(\theta)}\|X\lambda\| = \frac{\theta}{\sin(\theta)}\|X(X^TX)^{-1}X^Tx\|$. In conclusion the closest point of x in the subspace is:

$$y = X\lambda = \frac{X(X^TX)^{-1}X^Tx}{\|X(X^TX)^{-1}X^Tx\|} = \frac{\hat{x}}{\|\hat{x}\|}$$

4.2.2 Unexplained Variance and AUV criterion

In this section is described how to compute the *unexplained variance* and the *AUV criterion* described in Chapter 3.2.
Let's consider data points $\hat{y}_i$ living on the sphere, between them are chosen the reference ones. To compute the unexplained variance it is necessary to evaluate the projection y_i of all data points onto the subspace generated by the reference points and then the distance between the data-point and the projection is computed, $r_i = \text{dist}(\hat{y}_i, y_i)$. So the formula is the following:

$$\sigma^2_{out}(X) = \sum_i r_i^2(X) = \sum_i dist^2(\hat{y}_i, y_i(X))$$

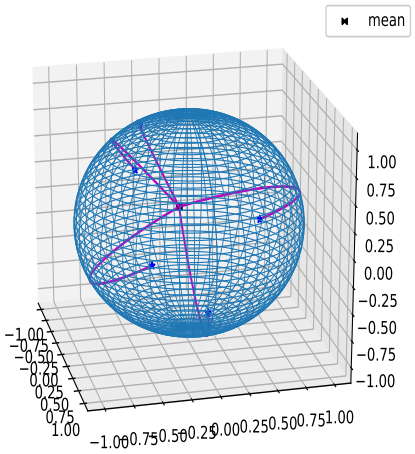


Figure 4.2: In this example there are Random points on the sphere and the mean is computed i.e. the point which minimize the sum of all the distance from the other points. This is a local maximum.

The function AUV (Accumulated Unexplained Variances described in 3.2) takes the first i-points of the reference matrix $X = [x_0, \cdots, x_i]$ and compute σ_{out} and sum it with all the previous $(i-1)$ σ_{out} computed before adding the i−th point. For example, the first step is minimize the distance of the data points with $X_0 = [x_0]$, then it is done the same with the subspace generated by $X_1 = [x_0, x_1]$ and sum up with the previous result (see figure 4.3). In general:

$$AUV = \sum_i \sigma_{out_i}(X_i, Data)$$

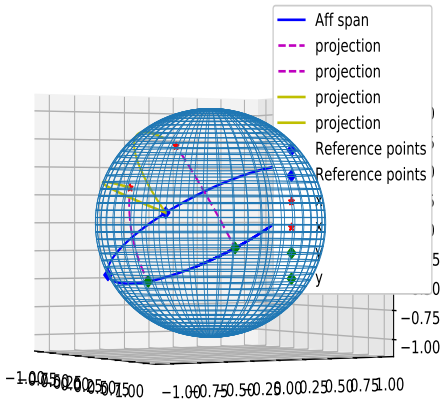


Figure 4.3: In this example we have two reference points and two data points, so the AUV function give the sum of the distance between the data points and the first reference point (yellow) and the distance between the data points and the subspace of the two reference points

4.3 Testing on data

In this section is tested how the three different kind of Barycentric Subspaces described in section 3 work on the 2−dimensional Sphere.
The data used below are taken from [15]:

Triangle	Pattern	w_1	w_2	w_3
1	RLR	0.48647	0.02354	−0.11310
2	LRL	0.48641	0.00490	0.11569
3	RLR	0.49371	−0.0050	−0.07889
4	LRL	0.49909	0.03003	0.00316
5	RLR	0.49672	−0.03147	−0.04777
6	LRL	0.46722, 0	02082	0.17684
7	RLR	0.47245	0.03951	−0.15882
8	LRL	0.48762	−0.07167	0.08422
9	RLR	0.49811	−0.03717	−0.02256

The first step is to normalized all the data, so each column is a points of

$\mathbb{R}^3$ on the sphere S^2:

$$\text{DataTrack} = \begin{pmatrix} 0.97294233 & 0.97281441 & 0.98742356 & 0.99817478 \\ 0.04708011 & 0.00979994 & -0.01000004 & 0.06005969 \\ -0.22620054 & 0.23137867 & -0.15778057 & 0.00631997 \end{pmatrix}$$

$$\begin{pmatrix} 0.99343386 & 0.93443928 & 0.94491165 & 0.97523442 & 0.99621176 \\ -0.06293961 & 0.04163997 & 0.07902097 & -0.14333918 & -0.07433939 \\ -0.09553941 & 0.35367973 & -0.31764392 & 0.16843904 & -0.04511963 \end{pmatrix}$$

The Data of the matrix Track are shown on the sphere in the figure below:

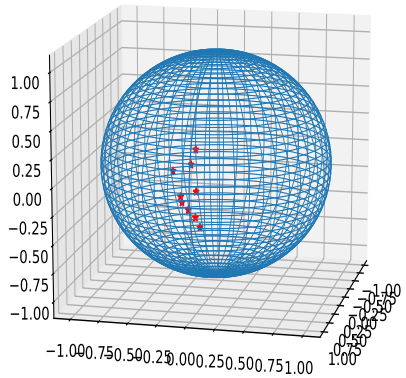


Figure 4.4: Track-Data points on the sphere

Firstly the unexplained variance, i.e. the Karcher mean of the Data is computed and the 1-Forward Barycentric Subspace (FBS), i.e. the geodesic passing through the mean and the second point which best fit the data, is shown. The results are shown in figure 4.5.

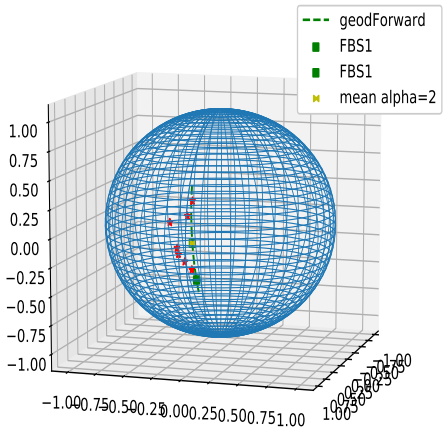


Figure 4.5: FBS

After Barycentric subspace (BSA) is computed. Looking at figure 4.6, it can be seen that BSA and FBS share the same geodesic, this means that the geodesic which minimizes AUV is also the one that pass to the mean and is built with FBS. Then both of these methods fit well the data.

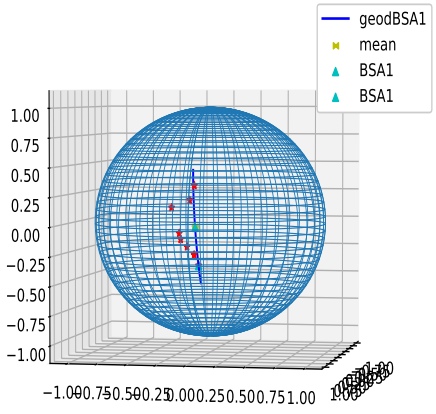


Figure 4.6: 1-BSA

In figure 4.7 it is shown the resulting points using the Pure Barycentric Subspace (PBS). The resulting points are different from FBS and BSA. In the figure is also underline the three points choose with 2-PBS to show that the previous points chosen with 1-PBS are not fixed.

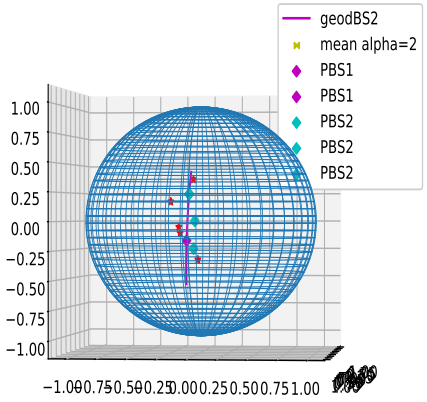


Figure 4.7: 1 and 2-Pure Barycentric subspace

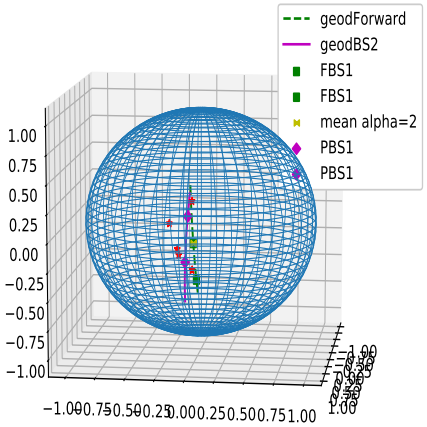


Figure 4.8: FBS and PBS

In conclusion the BSA and FBS gave the same results, while the PBS found different reference points to build the subspace. However the data are well fitted by all the subspaces.

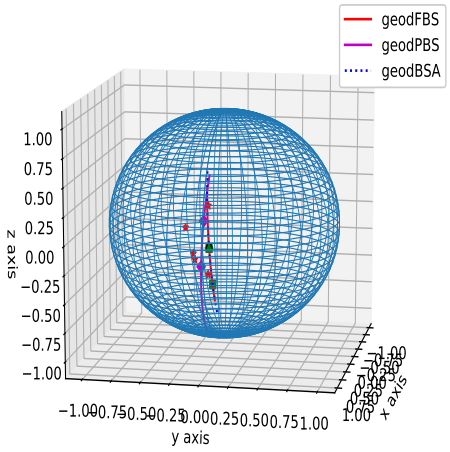


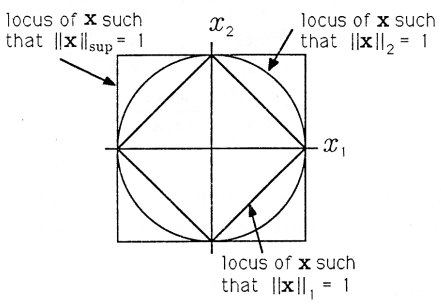
Figure 4.9: In this picture we have the Data track in red, the Mean in yellow. The geodesic for FBS are blue and are under the geodesic of BSA in light blue, while in magenta is the geodesic of PBS

4.3.1 Changing the norm

In the previews example and calculus it was used the 2–variance for computing the unexplained variance. Now it is interesting to study the difference if it is used the unexplained α–variance. As it was shown in section 3.1.1 this construction is very natural with barycentric subspaces since the affine span is stable under the choice of the value α.
Then the formula of $\alpha-variance$:

$$\sigma^{\alpha}_{out} = \sum_{i=0}^{k} \frac{1}{\alpha} \mathrm{dist}^{\alpha}(x, y)$$

The choice of α leads to a different influence of the points. Indeed different norms give different importance to the ouliers, data further from the other that can be, for example, corrupted by noise.
For example, in the unit circle: the 1-norm of it in $\mathbb{R}^2$ is a square, the 2-norm or the Eclidean norm is clearly the unit circle, while for the infinity norm it is a different square, and for any p-norm is a superellipse.



So changing the norm is changing the influences the points have on our study. The test is then compute with the same data of the previous section changing the value of $\alpha = 1$, $\alpha = 2$ and $\alpha = 0.1$.
The global minima of the $\alpha-$variance defines the Fréchet median for the value of $\alpha = 1$, the Fréchet mean for the classical $\alpha = 2$ and the barycenter of the support of the distribution if the value of $p = \infty$. Indeed the different solutions reach for the global minima, i.e. the first point of FBS, are computed and lead to different points of the data:

$$0\text{FBS}_2 = \begin{pmatrix} 0,99817478 \\ 0.06005969 \\ 0,00631997 \end{pmatrix} \quad 0\text{FBS}_1 = \begin{pmatrix} 0,99621176 \\ -0.07433939 \\ -0,04511963 \end{pmatrix} \quad 0\text{FBS}_{0.1} = \begin{pmatrix} 0,99343386 \\ -0.06293961 \\ -0,09553941 \end{pmatrix}$$

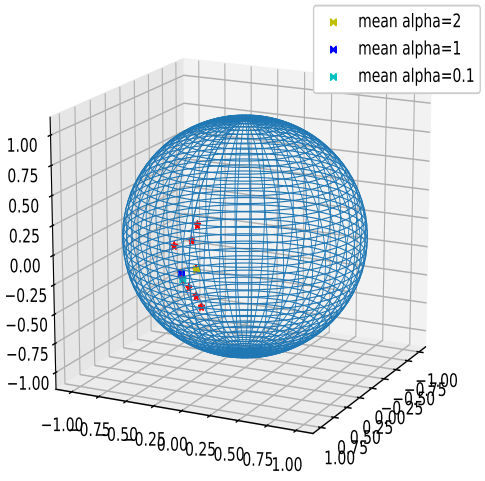


Figure 4.10: The first point of FBS with different values of alpha

The Forward barycentric subspace (L_α k–FBS) iteratively adds the point that minimizes the unexplained α–variance up to $k+1$ points, so the second step finding the geodesics lead to the following three different solutions:

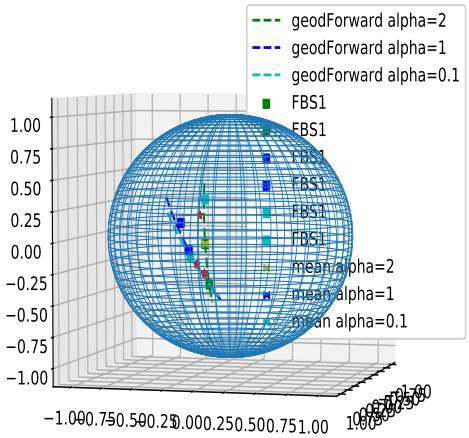


Figure 4.11: FBS different alpha

Then, the other two methods are computed, the optimal Pure Barycentric subspace with backward reordering (L_α k–PBS) estimates the $k+1$ points that minimize the unexplained α– variance, and then reorder the points accordingly for lower dimensions. The Barycentric Subspace Analysis of order k (L_α k–BSA) looks for the flag of affine spans defined by $k+1$ ordered points that optimized the L_α AUV.
All the results of PBS, FBS and BSA are then shown changing α:

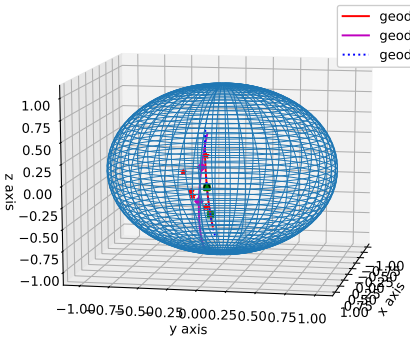


Figure 4.12: $\alpha = 2$

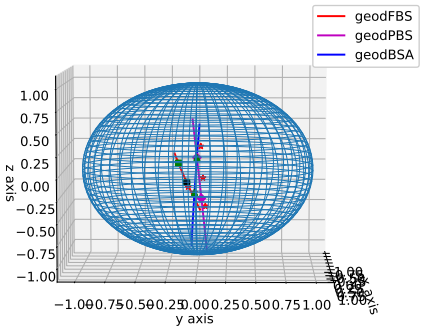


Figure 4.13: $\alpha = 1$

Figure 4.14: Final results

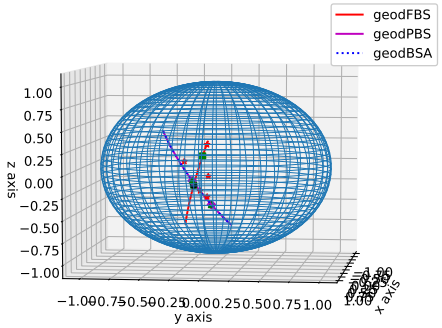


Figure 4.15: $\alpha = 0.1$

It can be noticed that, as seen before for $\alpha = 2$ BSA and FBS share the same geodesic while PBS not. For $\alpha = 1$ all the three subspaces are different, but PBS and BSA look to have geodesics more similar, the mean is contained only by FBS and this is the reason why it is a subspace more different from the other two. For $\alpha = 0.1$ BSA and PBS are equal while FBS is different, but the three subspaces contain the mean points as one of the reference points. On the other hand the forward method gives something less intuitive in term of robustness compared to the PBS and BSA.

4.3.2 Data points random distributed

In this section it is texted the behaviour of the different subspaces changing the norm and taking 100 random data points on the 2-dimensional Sphere.

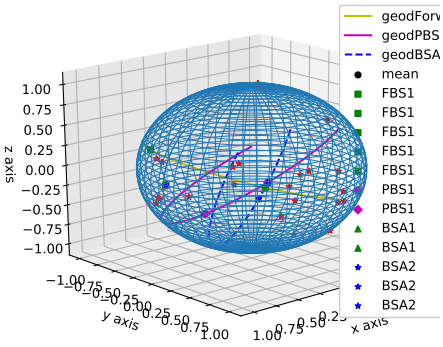


Figure 4.16: $\alpha = 2$

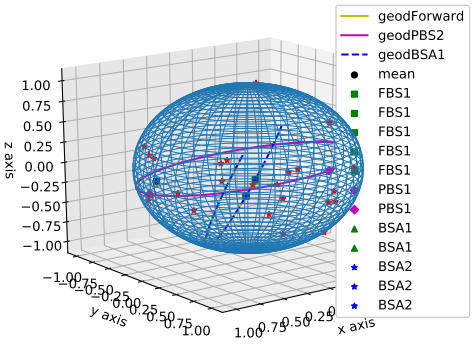


Figure 4.17: $\alpha = 1$

Figure 4.18: Final results

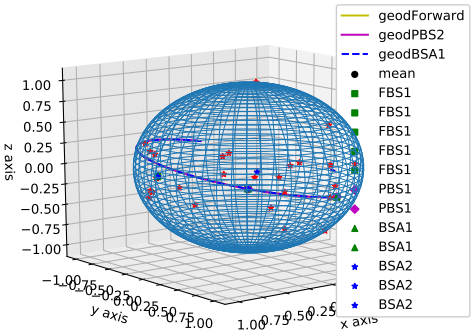


Figure 4.19: $\alpha = 0.1$

For $\alpha = 1$, FBS and BSA share the same geodesic. For $\alpha = 0.1$ all the three subspace have the same geodesic. For $\alpha = 2$ all the three subspaces have different geodesics.

4.3.3 Cluster on the Sphere

The last experiment is to see how the different Barycentric Subspaces behave in presence of clusters. Three clusters where plotted on the Sphere and the results are the following:

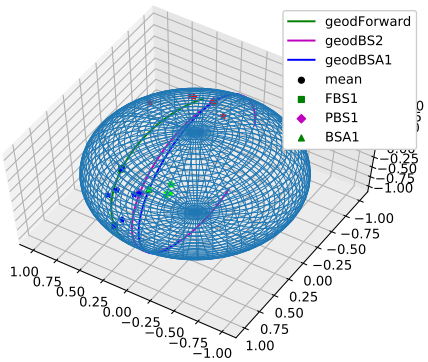


Figure 4.20: three cluster on the Sphere

4.4 Discussion on the different Barycentric Subspaces

FBS is computationally the less expensive because the previous points are fixed, while in PBS and BSA the points always change.
PBS probably best fit the data than FBS but the order given to the points at the end of the method is a big limit.
BSA takes care of the order just at the beginning (during the choice of the $k-$ points) but it is the most expensive.
In the first example FBS and BSA share the same geodesic while PBS not.
Changing the norm leads to different solution because the points are chosen thank to the weighted distance it is wanted to consider. Indeed we reach robustness to outlier data when $0 \leq \alpha \leq 1$. This can be observed also when we change norm in 100 hundred points.
In conclusion the fitting of the data is well-approximated by Barycentric Subspace Analysis, the generalization of AUV criterion taking care of the whole flag of subspaces is a really good criterion.

Chapter 5

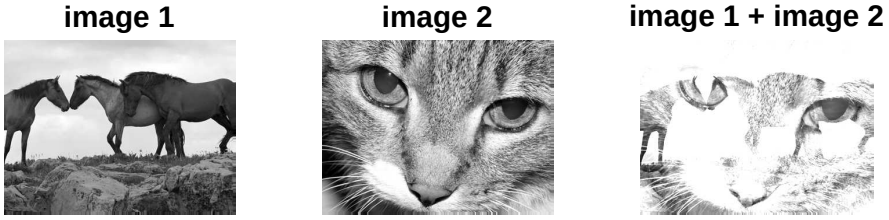
Barycentric Subspace applied on Images

The purpose of this Chapter is to show an interesting application of Barycentric Subspace on the problem of Image Registration. Image registration is a fundamental task in medical image processing, the goal of registration is to estimate the optimal transformation which maps an image to another. This process is used to align two images. There are several techniques to compute the registration, one of them consists in looking for a diffeomorphism which minimize the deformation between the two images, express via an energy function.

In the first section it is provided an introduction of the general registration methods and it is presented the *LCC Log-Demons Algorithm* for reference see [5] and [17]. The second part of the Chapter explains how use Barycentric Subspace to improve the registration with LCC Log-Demons algorithm for Cardiac Motion based on the work of M.M. Rohé [13].

5.1 Image Registration

The space of images cannot be described as a standard Euclidean space, since the sum of two images is not perceived as an image:



A general gray level image is expressed as a function defined on a 3D domain Ω, with values in $[0,1]$. Then a good approximation of the space of images $\mathcal{M}$ is a manifold where a point is an image $I \in \mathcal{M}$:

$$I : \Omega \subset \mathbb{R}^3 \to [0,1]$$

The goal of registration is to match the voxel intensities of the images in order to have a representation of the elements of the space. We will first present a registration procedure of an image I_1 with respect to a fixed image I_0. This will induced by a deformation which is a diffeomorphism $\Phi : A \to A$ such that $I_0 \approx I_1 \circ \Phi$. The formulation of the registration problem is then to find the diffeomorphism Φ which minimizes a functional of Energy, expressed as:

$$E(\Phi) = \frac{1}{\sigma_1^2}\mathrm{Sim}(I_0, I_1 \circ \Phi) + \frac{1}{\sigma_T^2}\mathrm{Reg}(\Phi) \tag{5.1}$$

where σ_1 is the noise image intensity parameter and σ_T is the parameter which controls the regularization term. The first term compute how similar are the images after the transformation is performed, since two images cannot

become perfectly equal and the simplest case is with L^2 norm $\mathrm{Sim}(I_0, I_1 \circ \Phi) = \|I_0 - I_1 \circ \Phi\|^2$. The second is a regularizing term of the function Φ, such that the transformation taken is as smaller and smoother as possible, and can be a Tikhonov regularizer.
This problem is in general ill-posed problem; for this reason an additive demons algorithm was proposed which adds a coupling term with a hidden variable c. Then the problem is to find the minimum of the global energy:

$$E(\Phi, c) = \frac{1}{\sigma_1^2}\mathrm{Sim}(I_0, I_1 \circ c) + \frac{1}{\sigma_x^2}\mathrm{dist}(\Phi, c)^2 + \frac{1}{\sigma_T^2}\mathrm{Reg}(\Phi)$$

with σ_x is the parameter for the spatial uncertain correspondences. In general the last two terms of the formula can have different forms depending on the problem. Note that the optimization problem give rise to two Euler Lagrangian equations, obtained by differentiating with respect to Φ and c respectively.
The classic demons algorithm introduces a *demons forces* u obtained from the decomposition: $c = \Phi + u$, which is the argument of the distance function d. The demons forces u are an effector situated in a point of the boundary object. Formally u is an increment from Φ to c, so that in manifold it has to be computed through an exponential map. In section 1.3.1 we pointed out that we can choose as distance $d(\Phi, c) = \|\delta v\|$.

5.1.1 Log-Demons Algorithm

The *Log Domain Algorithm* looks for the solution of the diffeomorphism Φ in the log-space. Indeed the group of Diffeomorphism can be seen as a Lie group structure, and this allows to work in its Lie algebra which is a vector space, so it is easier to work on compared to a manifold. We are not working on the whole infinite dimension group of diffeomorphisms but in a subset of finite dimension, but this is enough to our application. The exponential and its inverse logarithmic map, are local diffeomorphisms from the Lie algebra to the Lie group and vice versa. Then the problem is not looking directly to the diffeomorphism Φ but we are looking for the one parameter subgroup generated by a vector field lying in the tangent space $T_{Id}\mathrm{Diff}(\Omega)$, i.e. the Lie algebra. For a good compromise between the computational cost and the efficiency of the registration it has been proposed in [17] to work on the Stationary Velocity Field (SVF). If we arbitrary fix an origin in the space of diffeomorphisms, then the one parameter subgroup Φ of a SVF v is the unique solution of:

$$\frac{\partial \Phi(x, s)}{\partial s} = v(\Phi(x, s))$$

starting from the fixed origin. Since we are working in a group (see section 1.4.1), the solution is defined for $s \in [0,1]$, so that we can define $\exp(v) = \Phi(x,1)$. Then we can look for v in the log-space, which is the Lie algebra. The Log-Demons functional where $c = \exp(w)$ is:

$$E(v,w) = \frac{1}{\sigma_1^2}\mathrm{Sim}(I_0, I_1 \circ \exp(w)) + \frac{1}{\sigma_x^2}\mathrm{dist}(v,w)^2 + \frac{1}{\sigma_T^2}\mathrm{Reg}(v) \tag{5.2}$$

An important property of the exponential map is the non commutative property, so that the BCH formula (1.4.2) can be used. It has been proved that the first two terms of the decomposition provide already a good approximation if one of the terms is small, in our case: $\delta v = \mathrm{BCH}(-v,w) = \log(\exp(-v)\circ\exp(w))$.
This allows to minimize the functional with a two step algorithm:

1. *Optimization of Similarity* Given v:

   $$E(v,\delta v) = \frac{1}{\sigma_1^2}\mathrm{Sim}(I_0, I_1 \circ \exp(v) \circ \exp(\delta v)) + \frac{1}{\sigma_x^2}\|\delta v\|^2 + \frac{1}{\sigma_T^2}\mathrm{Reg}(v)$$

   The optimization is compute for δv, and thanks to BCH it is found $w = \mathrm{BCH}(v, \delta v)$

2. *Optimization of Regularization* Given w:

   $$E(v,w) = \frac{1}{\sigma_1^2}\mathrm{Sim}(I_0, I_1\circ\exp(w)) + \frac{1}{\sigma_x^2}\|\log(\exp(-v)\circ\exp(w))\|^2 + \frac{1}{\sigma_T^2}\mathrm{Reg}(v))$$

   optimized respect to v.

5.1.2 LCC-Log Demons Functional

In [5] was proposed to replace the classical similarity measure Sim of formula 5.1 with the local correlation coefficient (LCC), this is the metric used to measured the difference between the fixed image I_0 and the warped image I_1. The LCC is defined as $\mathrm{Sim}(I_0, I_1 \circ \Phi) = \rho(I_0, I_1 \circ \Phi)$, where

$$\rho(I_0, I_1 \circ \Phi) = \int_\Omega \frac{\overline{I_0 I_1}}{\sqrt{\overline{I_0^2}\,\overline{I_1^2}}}$$

where $\overline{I_0} = G_\sigma * I_0(x)$ is the Gaussian smoothing applied on the image. The coefficient varies between -1 (perfect negative correlation), 0 (no correlation) and 1 (perfect correlation). It measures how the intensities of the two images are correlated within a local neighborhood σ. Then the LCC similarity has been used with the Log-Demons algorithm to update $\delta v = w - v$.

5.2 Using Barycentric Subspace as a prior on the Registration

The methods proposed in [13] is to use Barycentric Subspace as a prior for the registration with respect to a set of images. Registration is usually done using a single image as a prior. This choice leads to compute really large deformation from the fixed image to other images. A good way to overcome this problem is to use a group-wise registration, which means to perform registration simultaneously by a group of images chosen as reference frames (see image 5.1). This incorporates the images information in registration process and eliminates bias towards a chosen reference frame. In this way we have to compute only small deformation, but it is needed to add a criteria to ensure consistency between different groups of images, since the registration of each image is computed with respect to different reference. To solve the problem instead of performing the registration with respect to the closest reference, it has been proposed by [13] to build a subspace containing these references and use it as a prior in the registration. So the main idea to improve the registration is to use barycentric subspace to build the subspace where we will perform the regularization term of formula 5.1.

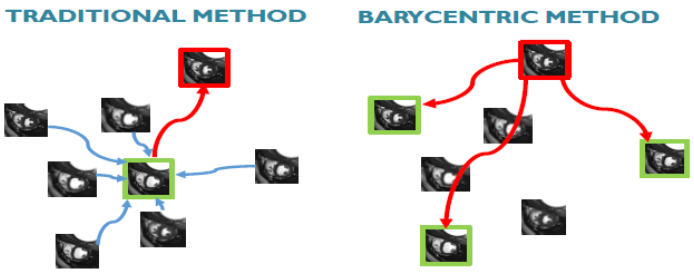


Figure 5.1: Image taken from [13]

Before presenting the registration with barycentric subspace, we need to explain how to find the projection of an image onto the barycentric subspace and how to choose the reference images between all the set og images.

5.2.1 Project an image onto the Barycentric Subspace

Let consider a set of N images $\{I_j\}_{j=1,\cdots,N}$. To build the Barycentric Subspace we need reference points and in this setting it means to select a set of reference images $\{R_i\}_{i=1,\cdots,k+1} \subset \{I_j\}_{j=1,\cdots,N}$.
Then the Barycentric Subspace is defined as the set of images $\hat{I}$ for which

there exists $(k+1)$ barycentric coefficients λ_j which satisfy the condition:

$$\sum_{j=1}^{k+1} \lambda_j \vec{\hat{I}R_j} = 0$$

where we have denoted $\vec{\hat{I}R_j}$ the logarithmic map. Using the stationary velocity fields (SVF) notation of section 5.1.1:

$$\sum_{j=1}^{k+1} \lambda_j v_j = 0$$

where v_j is the SVF mapping the image $\hat{I}$ to R_j.
As it was shown for the sphere in section 4.2.1 find the projection is a really crucial passage. So we want to compute the projection $\hat{I}$ of an image I within the subspace generated by $\{R_i\}_{i=1,\cdots,k+1}$ (see figure 5.2).

[H]

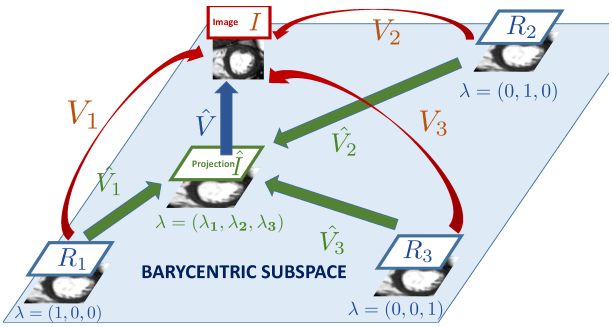


Figure 5.2: Barycentric subspace of dimension 2 built from 3 references images $(R_1; R_2; R_3)$. $\hat{I}$ is the projection of the image I within the barycentric subspace such that $\|\hat{v}\|$ is minimum under the conditions $\sum_j \lambda_j \hat{v_j}$ and $\hat{v} + \hat{v_j} = v_j$. Image from [13].

Following the computation of section 4.2.1, we want to minimize the distance between I and the projection $\hat{I}$ with constraint that the projection belongs to the Barycentric Subspace. The condition is expressed via a minimization of the norm of the SVF $\hat{v}$ which parametrized the deformation of I to $\hat{I}$. The request of belonging to the subspace is $\sum_j \lambda_j \hat{v_j}$, thank to BCH formula 1.4.2, we can use the first order approximation $v_i = \hat{v} + \hat{v}_i$. This means that the problem we want to solve is:

$$\min_{\hat{v}} \|\hat{v}\|^2 \quad \text{subject to} \quad \sum_i \lambda_i (v_i - \hat{v}) = 0$$

requiring that $\sum_i \lambda_i \neq 0$. It is not restrictive to add two other constraints on the Barycentric coefficients: first we ask $\sum_j \lambda_j = 1$, then they are normalized and second $\lambda_i \leq 1$ such that it forces the projection to lie within a border defined by the reference images. Then:

$$\min_{\hat{v}} \|\hat{v}\|^2 \quad \text{subject to} \quad \hat{v} = \sum_i \lambda_i v_i \quad \sum_j \lambda_j = 1 \quad \lambda_i \leq 1$$

Thus the Lagrangian becomes:

$$\Lambda(\boldsymbol{\lambda}, \alpha, \boldsymbol{\beta}) = \| \sum_i \lambda_i v_i \|^2 + \alpha(1 - \sum_i \lambda_i) + \boldsymbol{\beta}(1 - \boldsymbol{\lambda})$$

where $\boldsymbol{\lambda}, \boldsymbol{\beta}$ are vectors and α a scalar. Then deriving for λ_j:

$$\frac{\partial \Lambda(\boldsymbol{\lambda}, \alpha, \boldsymbol{\beta})}{\partial \lambda_j} = 0$$
$$2 < \sum_i \lambda_i v_i, v_j > -\alpha - k_j = 0 \quad \forall j$$

Deriving for α:

$$\frac{\partial \Lambda(\boldsymbol{\lambda}, \alpha, \boldsymbol{\beta})}{\partial \alpha} = 0$$
$$1 - \sum_j \lambda_j = 0 \quad \forall j$$

From the theory of Lagrangian multiplier with inequality constraints, the last term of the Lagrangian adds to the system two equations: $\beta_j \geq 0$ and $\beta_j(1 - \lambda_j) = 0$ for all j. So the solution can be found by solving the following system:

$$\begin{cases} < \sum_i \lambda_i v_i, v_j > = \frac{\alpha + k_j}{2} & \forall j \\ \sum_j \lambda_j = 1 & \forall j \\ \beta_j(1 - \lambda_j) = 0 & \forall j \\ \beta_j \geq 0 & \forall j \end{cases}$$

The analysis is divided in two cases if $\beta_j = 0$ or $(1 - \lambda_j) = 0$.
If all β_j are zero, then the inequality constraint is not filled and so we have to solve $< \sum_i \lambda_i v_i, v_j > = \frac{\alpha}{2}$ for all j. Denoting S the matrix of the scalar product $S_{ij} = < v_i, v_j >$:

$$S\boldsymbol{\lambda} = \alpha \mathbf{1}$$

Adding the condition: $\sum_i \lambda_i = 1$ for all j, the optimal barycentric coefficients $\boldsymbol{\lambda}^*$:

$$\boldsymbol{\lambda}^* = \frac{S^{-1}\mathbf{1}}{\sum_i (S^{-1}\mathbf{1})_i}$$

The second case is if some β_j are different from zero, then $\lambda_i = 1$ for these indices. Then we have to solve a lower dimensional problem removing the constrains already satisfied.
In conclusion, we have found the barycentric coordinates which allow us to find the SVF $\hat{v}$ just computing the weighted sum: $\hat{v} = \sum_i \lambda_i^* v_i$. Then the projected image $\hat{I}$ is given:

$$\hat{I} = I \circ \exp(\hat{v})$$

5.2.2 Choice of the reference images for cardiac motion

It is necessary to define a criterion to choose the reference images $\{R_i\}_{i=1,\cdots,k+1}$ within the set of images $\{I_j\}_{j=1,\cdots,N}$. The method proposed in [13] is to use an optimization approach. It consists on trying all the possible combination of $k+1$ references and choose the set of them that minimize the following functional:

$$(R_1, \cdots, R_{k+1}) = \arg\min \mathcal{E}(R_1, \cdots, R_{k+1}) = \arg\min \sum_j \|\hat{v}_j\|^2 \qquad (5.3)$$

where $\hat{v}_j$ is the projection of the image I_j to the barycentric subspace defined by $(R_1, \cdots, R_{k+1})$.

5.2.3 Barycentric Log-Demons Algorithm

Lastly we are going to explain the idea to improve the registration thanks to Barycentric Subspace. This is done using Barycentric Subspace as a prior. The goal is to find the registration of an image $I \in \{I_j\}_{j=1,\cdots,N}$ with respect to all the reference images R_j chosen between the set of images $\{I_j\}_{j=1,\cdots,N}$ using 5.3. The methodology of barycentric subspace is used in combination with LCC Log-Domain Diffeomorphic Demons algorithm described in section 5.1.2.
The first step is to initialize the algorithm with the standard registration of the image I with each references to get the SVF v_j^0. Using the projection onto the subspace explain in section 5.2.1 we can find the first approximation of the SVF $\hat{v}^0$ which encodes the distance between I and the subspace. Since we have v_j^0 and $\hat{v}^0$, we can find with BCH formula the SVF $\hat{v}_j^0$ which encodes the distance between the reference R_j and the projected image $\hat{I}$. This allows us to compute the warped references $\hat{R}_i^0 = R_i \circ \exp \hat{v}^0$.
Then the algorithm proceeds iterating the process, but using $\hat{R}_i$ in place of R_i. Introducing these warped references will allow to look for an increment in the deformation spaces, instead of the full deformation. In the iteration

we project the update *demons forces* (i.e. u) on the barycentric velocity until convergence, when $\hat{R}_i^j = R_i \circ \exp(\hat{v}_i^j)$ is under the fixed threshold.

At the end of the iteration we look for the SVF $\hat{v}$ by mapping each warped reference $\hat{R}_j^i$ to the current image and compose this SVF with the barycentric velocity to get an estimation of the full deformation: $v_i^j = \hat{v}_i^j + \hat{v}$. In this way we have found the deformation v_i^j mapping symmetrically each reference to the image.

5.2.4 Representation of Cardiac Motion using Barycentric Subspace

The method described in section 5.2.3 has been introduced in [13] to study cardiac motion. In the context of cardiac motion analysis, a cardiac cycle is described as a sequence of gray level images. A first step of the analysis is to find correspondences - the registration step - between each of the frame of the sequence. It allows to track the motion of the myocardium and compute deformation fields representing the motion during a cycle.
In a sequence of frames of cardiac motion, the regularization usually is done using one image (usually end diastole (ED)) as a prior. This choice leads to compute really large deformation from the ED to other images (for example to end systole (ES) image, see figure 5.3). Then the use of the methodology described is proposed.

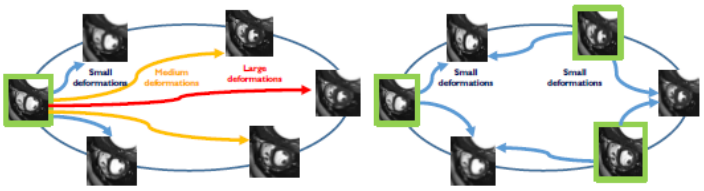


Figure 5.3: This image taken from [13] shows the difference of the classical registration approach for cardiac motion compared to the group-wise approach proposed with Barycentric Subspace.

The evaluation of the method is performed in [13] using one synthetic time serie of $T = 30$ cardiac image frames. The use of a synthetic sequence has the important advantage to provide a dense point correspondence field following the motion of the myocardium during the cardiac cycle which can be used to evaluate the accuracy of tracking. In figure 5.4 it is shown a comparison the barycentric approach with the standard approach where the registration between one of the reference and the current frame is done directly.
The barycentric method leads to a substantial reduction of the error (about

30%) in the largest deformations but at the cost of adding error for the small deformations evaluated at the frame near the respective references. The results of [13] are shown in figure 5.4 where it is seen how the better evaluation of large deformations improve the estimation of the volume curve.

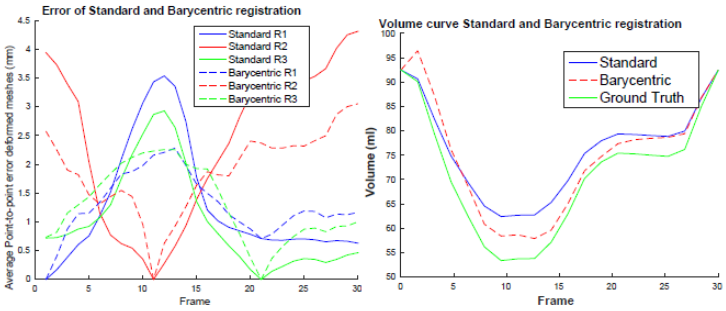


Figure 5.4: Image taken from [13]. (Left): Average point-to-point error on the meshes over the whole cycle between the ground-truth and the deformed meshes compared for the two methods (our proposed approach in dotted lines, and the standard one in plain lines). The registration is performed with respect to the three references for both cases. (Right): volume curves induced by the registration and comparison with the ground truth volume. Our proposed approach (red-dotted) performs a better approximation of the ground truth volume curve (green).

Chapter 6

Conclusion

In this thesis it was investigated the Barycentric Subspace Analysis. We presented different kind of barycentric subspaces depending on the choice of definition of mean, the three definitions lead to nested subspaces. We showed that the affine subspace in a Euclidean space can be generalized to manifolds thank to the definition of affine spans as the metric completion of Exponential Barycentric Subspace. This subspace is complete and this ensure that always exists a closest point on the subspace which is really important as we showed in the two example. We also saw that the definition of affine span is stable under the choice of different norm. Noticing that the affine subspaces can be easily nested as in the Euclidean case, it was presented the generalization of PCA: Barycentric Subspace Analysis.
Then we presented a deep studies on the Sphere showing that in this manifold the Barycentric Subspace are the great circle passing through the reference points. We showed how the three different methods: forward barycentric subspace, pure barycentric subspace and barycentric subspace analysis work on the Sphere. All this methods well-fit the data living on the Sphere, but the best fitting is the barycentric subspace analysis since it takes care of the whole flag of subspace using the AUV criterion.
Then we presented how the use of Barycentric Subspace can improve the registration in cardiac motion. It relies on building subspaces as the reference for registration instead of choosing a single image. So we presented how to choose the reference images, how to project an image inside the barycentric subspace and we concluded presenting the Barycentric Log-Demons Algorithm which performs and improves the registration step.

Appendix

Here, there is the Python code for the computation of the Riemannian setting on the Sphere and the Barycentric Subspaces.

```
#Sphere S_n embedded in Euclidean space R^{n+1}
def scal(x,y):
    molt=np.dot(x.T,y)
    return (molt)

#Random point on S_n
def RandomSph(n):
     #locals(x,xn)
     x=np.array([np.random.randint(-99,99,n+1)/100]).T
     nx=scal(x,x)
     while nx>1 or nx<1.e-6:
         x=np.array([np.random.randint(-99,99,n+1)/100]).T
         nx=scal(x,x)
     return(x/math.sqrt(nx))

########################################
# Spherical Exp and Log map distance  #
########################################

#Spherical Exponential
def ExpS(x,v,numeric):
    #locals(w,normw,a,b)
    #Project the vector v onto the tangent space TxS={w|scal
        (w,x)=0}
    w=v-scal(x,v)*x
    normw=math.sqrt(scal(w,w))
    # Numerical issues for theta.(y-cosh(theta). x)/sinh(
        theta)
    #if norm w close to zero we compute the Taylor expansion
    if numeric and N(normw)<1.e-6:
        a=1-(1/2)*normw**2+(1/24)*normw**4-(1/720)*normw
            **6+(1/40320)*normw**8
        b=1-(1/6)*normw**2+(1/120)*normw**4-(1/5040)*normw
            **6+(1/362880)*normw**8
    else:
```

```
        a=math.cos(normw)
        b=(math.sin(normw))/normw
    return(np.array((a*x+b*w), dtype='float'))

#################
# Spherical log #
#################

def LogS(x,y,numeric):
    #if scal(x,x)<>1 then return "x is not spherical" end if
    #if scal(y,y)<>1 then return "y not spherical" end if
    theta=math.acos(scal(x,y)[0,0])
    #Numerical issues for theta.(y-cos(theta). x)/sin(theta)
    #if theta**2 is small we compute the Taylor expansion
    if numeric and N(theta**2)<1.e-12:
        a=1+(1/6)*theta**2+(7/360)*theta**4+(31/15120)*theta
            **6+(127/604800)*theta**8
        b= 1-(1/3)*theta**2-(1/45)*theta**4-(2/945)*theta
            **6-(1/4725)*theta**8
    else:
        a=theta/math.sin(theta)
        b=theta*cot(theta)
    return(np.array((a*y-b*x), dtype='float'))

########################
# Riemannian Distance #
########################

def DistS(x,y):
    ##scal must belong to 0 and p
    if scal(x,y)>1:
        return(np.arccos(1.))
    return(np.arccos(scal(x,y))) #Riemannian distance on S^n
         inherited from the euclidean metric

##############################
## Barycentric Subspaces    ##
##############################

def RandomXRef(n,k):
    ##generate k+1 random points on S^n
    X=np.zeros((n+1,k+1))
    for i in range(k+1):
      X[:,i]= RandomSph(n).T
    return(X)

def projectionBS(x,X):
    ###Closest point of x onto Aff(X)
    n=size(x)-1
```

```
    Omega=np.dot(X.T,X)
    Omega=np.linalg.pinv(Omega)
    #Unconstrained projection on Aff(X)
    z=np.dot(np.dot(np.dot(X,Omega),X.T),x)
    return(z/math.sqrt(scal(x,z)))

def CoordOnBS(x,X):
    ###Coordinates on Aff(X)
    Omega=np.dot(X.T,X)
    # Unconstrained projection on Aff(X)
    w=np.dot(np.dot(np.linalg.pinv(Omega),X.T),x)
    return(w/math.sqrt(scal(x,np.dot(X,w))))

################################
##   alpha change the norm    ##
################################

def sigma_out(X,Data,alpha):
    ##sigma_out(X)=Sum(Residue**2(X))=Sum(dist**2(x,y))
    sig=0.
    for i in range(size(Data[0,:])):
        x=np.array([Data[:,i]]).T
        y=projectionBS(x,X)
        res=(abs(DistS(x,y)))**alpha/alpha
        sig=sig+res
    return(sig/(size(Data[0,:])))

def AUC(X,Data,alpha):
    ##AUC=Sum(sigma_out(Xi,Data))
    AUCtmp=0
    for i in range(1,size(X[0,:])+1):
        #select firts i points
        Xi=X[:,0:i]#in maple Xi := X[1..-1,1..i]; I take the
            first i points of X
        sigi=sigma_out(Xi,Data,alpha)
        AUCtmp=AUCtmp+sigi
    return(AUCtmp)

################################
## Best point approx of data ##
################################

def DataMean(Data,alpha):
    ### initialize with first point
    x=np.array([Data[:,0]]).T
    sigmin=sigma_out(x,Data,alpha)
    for i in range(1,size(Data[0,:])):
        sig=sigma_out(np.array([Data[:,i]]).T,Data,alpha)
        if N(sig<sigmin):
```

```
            sigmin=sig
            x=np.array([Data[:,i]]).T
    return(x,N(sigmin))

#########
## FBS ##
#########

##Compute next mode
def ForwardBS(Xin,Data,alpha):
    ##Compute the Forward Barycentric Subspace: minimizing
        sigma_out
    X=Xin
    sigmin=sigma_out(X,Data,alpha)
    for i in range(size(Data[0,:])):
        Xtest=np.concatenate((Xin,np.array([Data[:,i]]).T),
            axis=1)
        sig=sigma_out(Xtest,Data,alpha)
        if N(sig<sigmin):
            sigmin=sig
            X=Xtest
    return(X,N(sigmin))

#########
## PBS ##
#########

def SwapCol(X,i,j):
    ##Function that swap the column i and j of a matrix X
    col=np.copy(X[:,i])
    X[:,i]=X[:,j]
    X[:,j]=col
    return(X)

def Compute_PBS1(Data,alpha):
    ## 1-Pure Barycentric subspace = optimal geodesic
    X=Data[:,0:2]#Take column 0 and column 1
    sigmin=sigma_out(X, Data,alpha)
    for i in range(size(Data[0,:])):
        for j in range(i+1,size(Data[0,:])):
            Xtest=np.concatenate((np.array([Data[:,i]]).T,np.
                array([Data[:,j]]).T),axis=1)
            sig=sigma_out(Xtest,Data,alpha)
            if N(sig<sigmin):
                sigmin=sig
                X=Xtest
    ## Reorder the 2 points
    if N(sigma_out(np.array([X[:,0]]).T,Data,alpha))>N(
        sigma_out(np.array([X[:,1]]).T,Data,alpha)):
```

```
        SwapCol(X,0,1)
    return(X,N(sigmin))

def Compute_PBS2(Data,alpha):
    ## 2-Pure Barycentric subspace
    X=Data[:,0:3]
    sigmin=sigma_out(X, Data,alpha)
    for i in range(size(Data[0,:])):
        for j in range(i+1,size(Data[0,:])):
            for k in range(j+1,size(Data[0,:])):
                Xtest=np.concatenate((np.array([Data[:,i]]).T,
                    np.array([Data[:,j]]).T,np.array([Data[:,k
                    ]]).T),axis=1)
                sig=sigma_out(Xtest,Data,alpha)
                if N(sig<sigmin):
                    sigmin=sig
                    X=Xtest
        # Reorder the 3 points
        # find minimum
    sigmin=N(sigma_out(np.array([X[:,0]]).T,Data,alpha))
    imin=0
    for i in range(1,3):
        sig=N(sigma_out(np.array([X[:,i]]).T,Data,alpha))
        if sig<sigmin:
            imin=i
            sigmin=sig
    X=SwapCol(X,0,imin)#0 or 1??
    Xtest=np.concatenate((np.array([X[:,0]]).T,np.array([X
        [:,2]]).T),axis=1)
    if N(sigma_out(X[:,0:2],Data,alpha))>N(sigma_out(Xtest,
        Data,alpha)):
          X=SwapCol(X,1,2)##  X := < X[1..-1,1] | X[1..-1,3]
             | X[1..-1,2]
    return(X,N(sigma_out(X,Data,alpha)))

def Compute_PBS3(Data,alpha):
    ## 3-Pure Barycentric subspace
    X=Data[:,0:4]
    sigmin=sigma_out(X,Data,alpha)
    for i in range(size(Data[0,:])):
        for j in range(i+1,size(Data[0,:])):
            for k in range(j+1,size(Data[0,:])):
                for l in range(k+1,size(Data[0,:])):
                    Xtest=np.concatenate((np.array([Data[:,i]]).
                        T,np.array([Data[:,j]]).T,np.array([Data
                        [:,k]]).T,np.array([Data[:,l]]).T),axis
                        =1)
                    sig=sigma_out(Xtest,Data,alpha)
                    if N(sig<sigmin):
```

```
                    sigmin=sig
                    X=Xtest
    ## Reorder the 4 points
    ## find minimum
    sigmin=sigma_out(np.array([X[:,0]]).T,Data,alpha)
    imin=0
    for i in range(1,4):
        sig=N(sigma_out(np.array([X[:,i]]).T,Data,alpha))
        if sig<sigmin:
            imin=i
            sigmin=sig
    #exchange point 1 with imin
    SwapCol(X,0,imin)
    #Now repeat for the minimum pair
    imin=1
    sigmin=N(sigma_out(X[:,0:2],Data,alpha))
    for i in range(2,4):
        sig=N(sigma_out(np.concatenate((np.array([X[:,0]]).T,
            np.array([X[:,i]]).T),axis=1),Data,alpha))
        if sig<sigmin:
            imin=i
            sigmin=sig
    #Exchange X[1] with X[imin]
    if imin>2:
        SwapCol(X,1,imin)
    #last level
    if N(sigma_out(np.concatenate((np.array(X[:,0:2]),np.
        array([X[:,3]]).T),axis=1),Data,alpha))<N(sigma_out(
        np.array(X[:,0:3]),Data,alpha)):
        SwapCol(X,2,3)
    return(X,N(sigma_out(X,Data,alpha)))

#########
## BSA ##
#########

def Compute_BSA1(Data,alpha):
    ## 1-BSA # with less calculus: time divided by 6
    ## minimizing AUC
    X=Data[:,0:2]
    aucmin=AUC(X,Data,alpha)
    for i in range(size(Data[0,:])):
        X1=np.array([Data[:,i]]).T
        sig1=sigma_out(X1,Data,alpha)
        if sig1<aucmin:# otherwise sig1 + something cannot be
             smaller than aucmin
            for j in range(size(Data[0,:])):
                Xtest=np.concatenate((X1,np.array([Data[:,j]]).
                    T),axis=1)
```

```
            auc=sig1+sigma_out(Xtest,Data,alpha)
            if N(auc<aucmin):
                aucmin=auc
                X=Xtest
    return(X,N(aucmin))

def Compute_BSA2(Data,alpha):
    ## 2-BSA # time divided by 20! (4.3 instead of 85 on
        Tracks)
    X=Data[:,0:3]
    aucmin=AUC(X,Data,alpha)
    for i in range(size(Data[0,:])):
        X1=np.array([Data[:,i]]).T
        auc1=sigma_out(X1,Data,alpha)
        if auc1<aucmin:# otherwise auc1 + something cannot be
            smaller than aucmin
            for j in range(size(Data[0,:])):
                X2=np.concatenate((X1,np.array([Data[:,j]]).T),
                    axis=1)
                auc2=auc1+sigma_out(X2,Data,alpha)
                if auc2<aucmin:# otherwise auc2 + something
                    cannot be smaller than aucmin
                    for k in range(size(Data[0,:])):
                        X3=np.concatenate((X2,np.array([Data[:,k
                            ]]).T),axis=1)
                        auc3=auc2+sigma_out(X3,Data,alpha)
                        if N(auc3<aucmin):
                            aucmin=auc3
                            X=X3
    return(X,N(aucmin))

def Compute_BSA3(Data,alpha):
    ## 3-BSA
    X=Data[:,0:4]
    aucmin=AUC(X,Data,alpha)
    for i in range(size(Data[0,:])):
        #print('Iteration /n',i)
        X1=np.array([Data[:,i]]).T
        auc1=sigma_out(X1,Data,alpha)
        if auc1<aucmin:# otherwise auc1 + something cannot be
            smaller than aucmin
            for j in range(size(Data[0,:])):
                X2=np.concatenate((X1,np.array([Data[:,j]]).T),
                    axis=1)
                auc2=auc1+sigma_out(X2,Data,alpha)
                if auc2<aucmin:## otherwise auc2 + something
                    cannot be smaller than aucmin
                    for k in range(size(Data[0,:])):
```

```
                    X3=np.concatenate((X2,np.array([Data[:,k
                        ]]).T),axis=1)
                    auc3=auc2+sigma_out(X3,Data,alpha)
                    if auc3<aucmin:
                        for l in range(size(Data[0,:])):
                            X4=np.concatenate((X3,np.array([
                                Data[:,l]]).T),axis=1)
                            auc4=auc3+sigma_out(X4,Data,alpha)
                            if N(auc4<aucmin):
                                aucmin=auc4
                                X=X4
    return(X,N(aucmin))
```

Bibliography

[1] Bauer, M., Bruveris, M., Michor, P.W. *Uniqueness of the Fisher-Rao metric on the space of smooth densities.*

[2] Joshi, S., Srivastava, A., Jermyn, I.H. (2007) *Riemannian analysis of probability density functions with applications in vision.*, in 2007 IEEE Conference on Computer Vision and Pattern Recognition ; proceedings. Piscataway, NJ: IEEE, pp. 1664-1671.

[3] Joshi, S., Davis, B., Jomier, M., Gerig, G., *Unbiased diffeomorphic atlas construction for computational anatomy*, NeuroImage, Elsevier, 2004.

[4] Jost, J. *Riemannian Geometry and Geometric analysis*, Springer, 2008, fifth edition.

[5] Lorenzi, M., Ayache, N., Frisoni, G., Pennec, X. *LCC-Demons: a robust and accurate symmetric diffeomorphic registration algorithm.* NeuroImage, Elsevier, 2013, 81 (1), pp. 470-483.

[6] Lorenzi, M., Pennec, X., *Geodesics, Parallel Transport, One-parameter Subgroups for Diffeomorphic Image Registration* International Journal of Computer Vision, Springer Verlag, 2013, International Journal of Computer Vision, 105(2), pp. 111-127.

[7] Pennec, X. *Barycentric Subspace Analysis on Manifold*, Submitted to the Annals of Statistics, imsart-aos ver. 2017 .

[8] Pennec, X., Arsigny, V. *Exponential Barycenters of the Canonical Cartan Connection and Invariant Means on Lie Groups.* Frederic Barbaresco and Amit Mishra and Frank Nielsen. Matrix Information Geometry, Springer, pp. 123-168, 2012, 978-3-642-30231.

[9] Pennec, X. *From Riemannian Geometry to Computational Anatomy*, Asclepios project-team, INRIA Sophia-Antipolis Mediterranee, 2011.

[10] Pennec, X. *Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements.* Journal of Mathematical Imaging and Vision, Springer Verlag, 2006, 25(1), pp.127-154.

[11] Pennec, X. *Sample-limited L_p Barycentric Subspace Analysis on Constant Curvature Spaces* Nov 2017, Paris, France. Springer International Publishing, Lecture Notes in Computer Science (LNCS), pp.20-28, 2017, Geometric Science of Information.

[12] Rohé, M.M., Sermesant, M., Pennec, X. *Barycentric Subspace Analysis: a new Symmetric Group-wise Paradigm for Cardiac Motion Tracking.* MICCAI 2016-the 19th International Conference on Medical Image Computing and Cmputer Assisted Intervention, Oct 2016, Athens, Greece 9902 (448-456), pp. 300-307,2016 MICCAI 2016, Lecture Notes in Computer Science.

[13] Rohé, M.M, Sermesant, M., Pennec,X *Low-Dimensional Representation of Cardiac Motion Using Barycentric Subspaces: a New Group-Wise Paradigm for Estimation, Analysis, and Reconstruction.*

[14] Rohé, M.M. *Reduced representation of segmentation and tracking in cardiac images for group-wise longitudinal analysis* PhD Thesis, Université Côte d'Azur, July 2017.

[15] Small, Christopher G. *The Statistical Theory of Shape*, Springer series in statistics, 1996.

[16] Tu, L.W. *An introduction to Manifolds*, Springer, 2011, second edition.

[17] Vercauteren T., Pennec X., Perchant A., Ayache N.,2008 *Symmetric log-domain diffeomorphic registration: A demons-based approach* Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI'08), Part I, Springer, New York, USA, pp. 754-761.

[18] Warner F. *Fundamentation of differentiable maanifold*, University of Pennsylvania, Scott Foresman and Company.

[19] Wilkins, D. R. *A Course in Riemannian Geometry*, 2005.