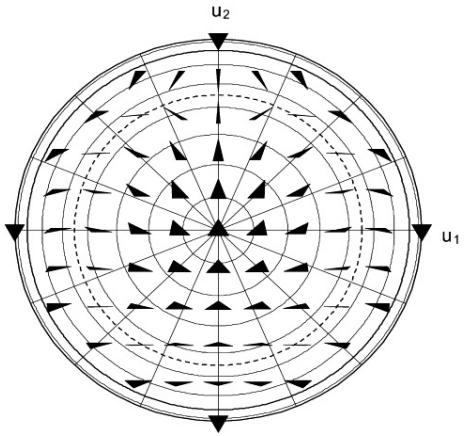
Statistics on Manifolds applied to Shape Theory

Author : Matthieu Simeoni
Supervisor : Pr. Victor Panaretos

June 26, 2013



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Contents

1. Introduction to Shape Theory
2. Introduction to differential manifolds
3. Principal Components Analysis
4. Nonlinear statistics on manifolds
5. Shape space of triads
6. Conclusion

Introduction to Shape Theory

We'd like to propose a **mathematical definition of shape of objects** and design **statistical tools to analyze** those shapes.

Definition

The shape of an object can be define as the total of all information that is invariant under translations, rotations and rescaling (invariance under similarity transformations).

Roughly speaking, we remove all information concerning location, scale and orientation. The obtained space is generally a differential manifold, called space of shapes.
Manifolds can have non null curvature, so we have to re-design all our classic linear statistics tools (means, PCA...).

In practice (Kendall school)

In shape theory we focus on special points called landmarks : they are points of special interest for the considered object, which are meant to provide a partial geometric description of it (see fig. 1).

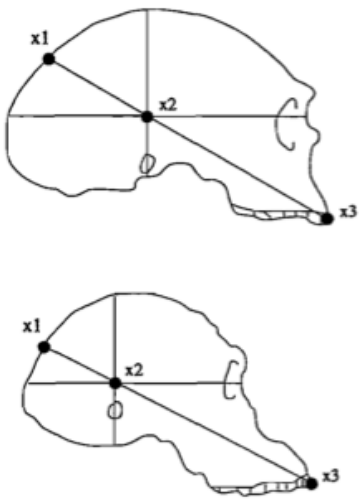


Figure: *Landmarks of human skulls (Neanderthal and australopithecine).*

Topological and differential manifolds

Let M^p be a **topological space** and $\{U_\alpha\}_{\alpha \in A}$ s.t $\cup_\alpha U_\alpha = M^p$. We also assume the existence of functions :

$$c_\alpha : U_\alpha \to \mathbb{R}^p,$$

that are all homeomorphisms onto the open subsets $c_\alpha(U_\alpha) \subset \mathbb{R}^p$.

Definition (Charts)

We say that the functions c_α are **charts** on M^p provided that :

$$c_\beta \circ c_\alpha^{-1} : c_\alpha(U_\alpha \cap U_\beta) \to c_\beta(U_\alpha \cap U_\beta), \quad (1)$$

is a **homeomorphism** from $c_\alpha(U_\alpha \cap U_\beta)$ to $c_\beta(U_\alpha \cap U_\beta)$, $\forall \alpha, \beta \in A$.

Definition (Atlas and topological manifold)

The collection $\{(U_\alpha, c_\alpha)\}_{\alpha \in A}$ forms an atlas on M^p. The set M^p together with its atlas is called a topological manifold of dimension p. If $c_\beta \circ c_\alpha^{-1}$ are $\mathcal{C}^r$-diffeormorphisms then M^p is said to be a $\mathcal{C}^r$-differential manifold.

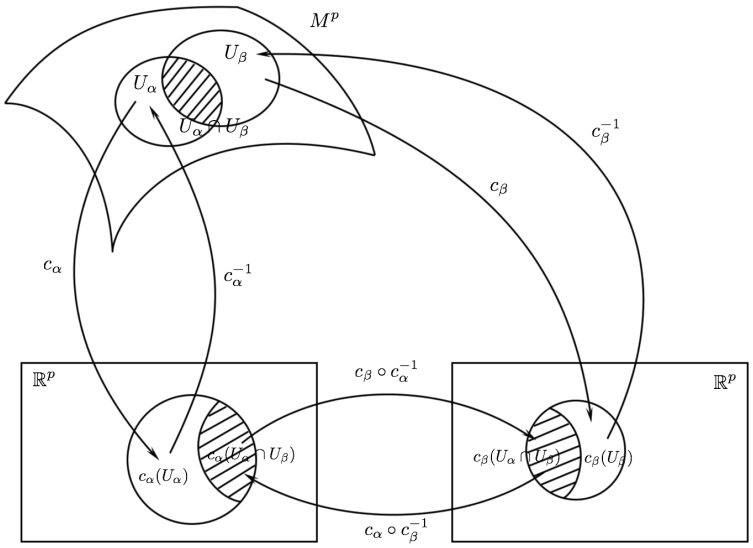


Figure: *Charts provide local coordinate systems on M^p. The patching criterion eq. (1) ensures the compatibility of two coordinate systems on a region of overlapping.*

Tangent vectors and tangent spaces

We define tangent vectors through equivalence classes of paths onto the manifold : *two paths $x(t)$, $y(t)$ passing through $x_0 \in M^p$ at $t = 0$ are said to be equivalent if they are tangent in x_0.*

Definition (Tangent vectors)

We define the tangent vector $\dot{x}$ to the path $x(t)$ at the point $x_0 = x(0)$ to be the equivalence class of $x(t)$ under the above equivalence relationship.

Definition (Tangent space)

The set of all tangent vectors to the manifold M^p at x_0 is called tangent space at x_0 and is denoted by $T_{x_0}(M^p)$.

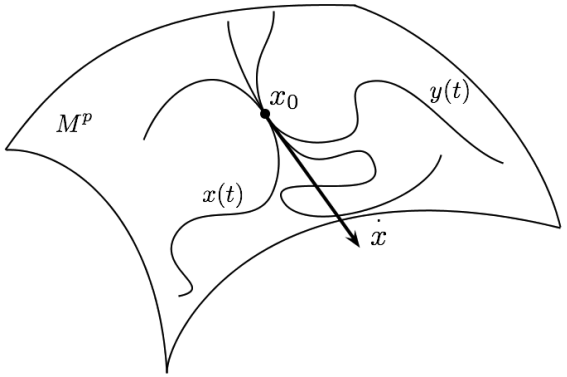


Figure: *Tangent vector $\dot{x}$ at x_0, seen as the equivalence class of all smooth paths tangent in x_0.*

We can provide a linear structure to our tangent space, by defining an addition and scalar multiplication on the above tangent vectors. We will exploit this linear structure to re-design our classic statistic tools for manifolds.

Geodesics

Provided the existence of a **metric tensor** $\forall x \in M^p$, one can define an inner product on $T_x(M^p)$. This metric structure on the tangent space allow us to define the length of a path $x(t)$:

$$L = \int_{t_0}^{t_1} \|\dot{x}(t)\| dt.$$

Definition (Geodesics)

A geodesic, is a smooth path $x(t)$ in a manifold which is locally the shortest.

Euler-Lagrange equations (characterization of geodesics) :

$$\frac{\partial F}{\partial \gamma_i}(t, \gamma_i, \dot{\gamma}_i) - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{\gamma}_i}(t, \gamma_i, \dot{\gamma}_i)\right) = 0, \quad (2)$$

with $F(t, \gamma_i, \dot{\gamma}_i) = \|\dot{\gamma}(t)\|$.

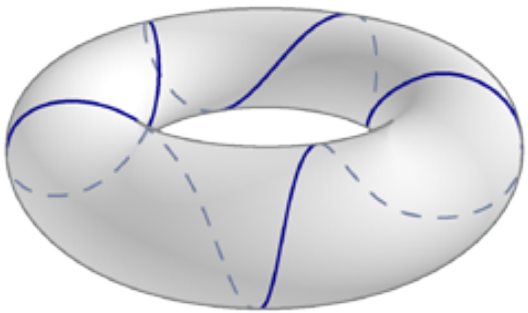


Figure: *Example of a geodesic path on a torus.*

The above equation can be re-written as a second order ODE, ensuring existence and unicity of geodesics with prescribed initial conditions (Picard-Lindelöf theorem).

The Exponential Map

We now wish to define a map from the tangent plane to the manifold.

Definition (Exponential Map)

Let M^p be a Riemannian manifold, $x \in M^p$, $v \in T_x(M^p)$ and $\gamma_v(t)$ the unique geodesic such that $\gamma(0) = x$ and $\dot{\gamma}(0) = v$. Then, we define the **exponential map** as :

$$\mathrm{Exp}_x(v) = \gamma_v(1).$$

- We have : $\mathrm{Exp}_x(tv) = \gamma_{tv}(1) = \gamma_v(t)$
- The exponential map is a local diffeomorphism between $T_x(M^p)$ and M^p (local inverse theorem).
- In the **injectivity radius** , we call the inverse of the exponential map the logarithmic map that we shall note Log_x.
- Computing geodesic distance :

$$d(x, \mathrm{Exp}_x(v)) = \|v\| \Rightarrow d(x, y) = \|\mathrm{Log}_x(y)\|$$

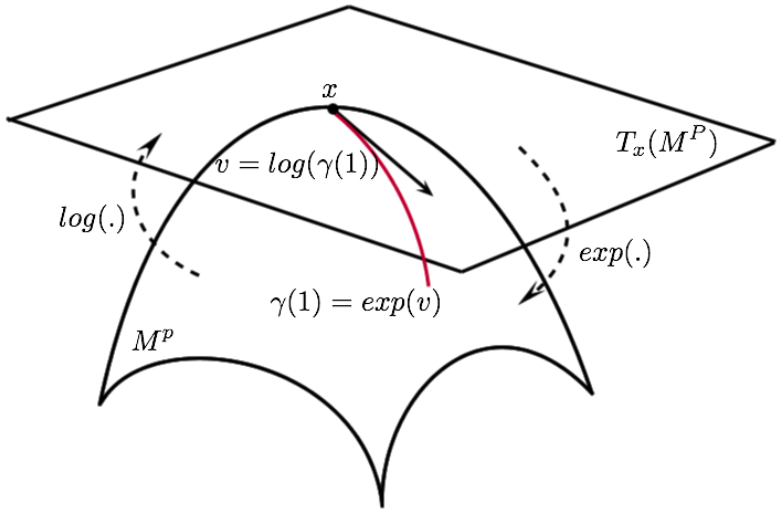


Figure: *The exponential and logarithm maps between a differential manifold ant the tangent plane at one point.*

Principal Components Analysis

PCA is a purely descriptive technique, aiming to re-express the data in an *optimal* way, as a sum of "independent" variables (same philosophy as Fourier Analysis). This allow to perform efficient exploratory analysis on large data sets, by reducing dimensionality.

Definition (PCA)

Let $X \in \mathbb{R}^p$ a random vector with known covariance matrix Σ. We define the p **principal components** as :

- First principal component :

$$w_1 = \underset{\|w\|=1}{\operatorname{argmax}} \operatorname{Var}(w^T X).$$

- k-th principal component, $2 \leq k \leq p$:

$$w_k = \underset{\|w\|=1}{\operatorname{argmax}} \operatorname{Var}(w^T \hat{X}_k),$$

with $\hat{X}_k = X - \sum_{i=1}^{k-1} ({w_i}^T X) w_i$.

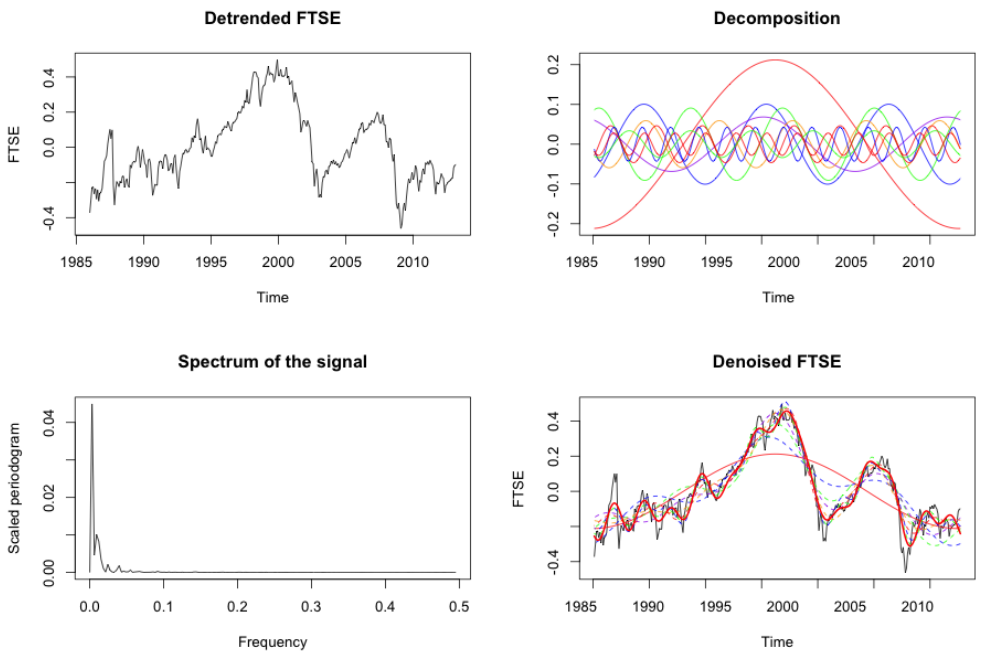


Figure: *Fourier decomposition of the FTSE from 1986 to nowadays. As PCA, Fourier Analysis aims to re-express the data, but this time as sum of independents trigonometric functions.*

PCA and Spectral Decomposition

We can show that the obtain w_i are the eigenvectors of Σ. De-correlating the data is the same as diagonalizing Σ.

Proposition (Ellipsoids and Principal Components)

Let X and Σ as above. Consider the family of p-**dimensional ellipsoids** :

$$X^T \Sigma^{-1} X = c, \qquad (3)$$

with c a constant. Then, the principal components define the directions of the principal axes of these ellipsoids.

The ellipsoids can be used as a measurement of the dispersion in term of variance of the data around the mean.

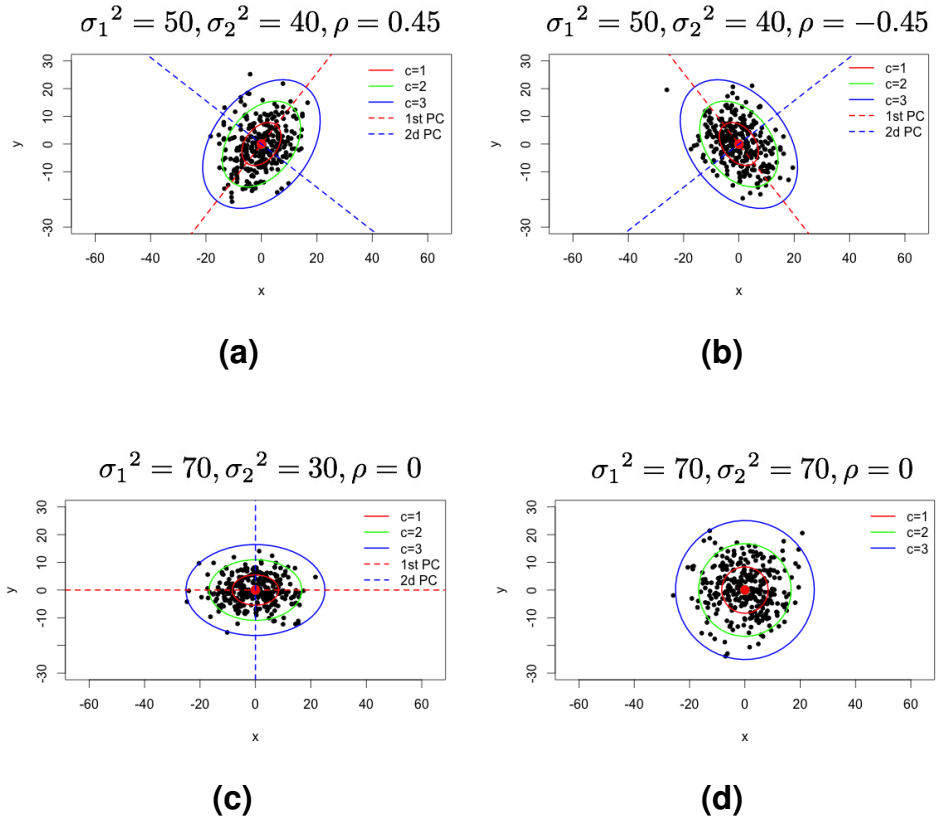
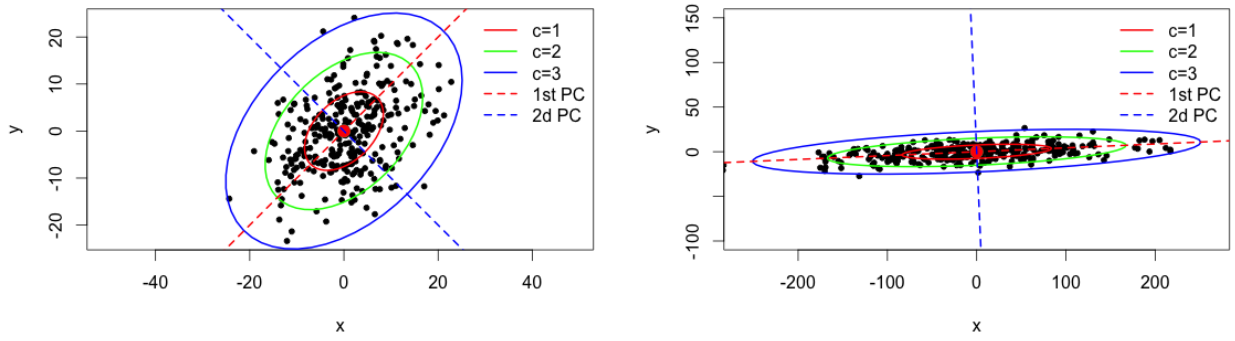


Figure: *Dispersion ellipsoids for a bivariate normal distribution with mean* $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ *and covariance matrix* $\Sigma = \begin{pmatrix} \sigma_1{}^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2{}^2 \end{pmatrix}$.

PCA sensitivity

- PCA can be very sensitive to units of measurement, so it could be wise to use the correlation matrix (covariance matrix of the standardized version of X) instead of the covariance matrix.



(a) *x measured in centimeters* **(b)** *x measured in millimeters*

Figure: *Sensitivity of the PCs and the dispersion ellipsoids to the units of measurement.*

- When we don't know Σ, we must use a proxy of it : the sample covariance matrix (or sample correlation matrix to avoid sensibility issues).

Means on Manifolds

Given $x_1, \ldots, x_N \in \mathbb{R}^d$, the **mean** $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is minimizing : $\bar{x} = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{N} \|x - x_i\|^2$.

This inspires the following extension :

Definition (Extrinsic mean)

For every $x_1, \ldots, x_N \in M^p$, we define the **extrinsic mean** as :

$$\mu_\Phi = \underset{x \in M^p}{\operatorname{argmin}} \sum_{i=1}^{N} \|\Phi(x) - \Phi(x_i)\|^2,$$

with $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^p$, and $\Phi : M^p \to \mathbb{R}^d$ an embedding.

- Can be computed with a gradient descent algorithm.
- Computationally convenient but extrinsic definition (requires an embedding).

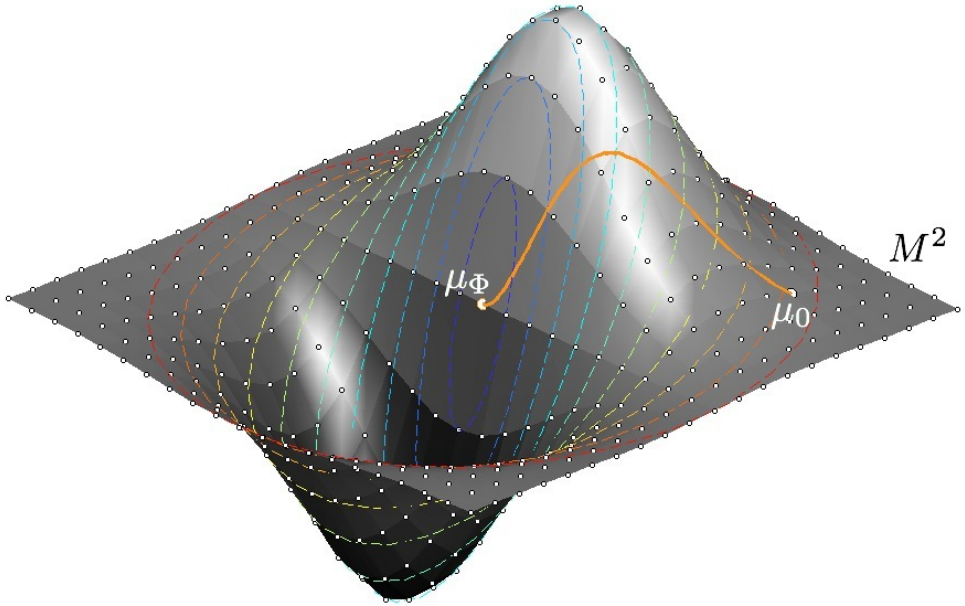


Figure: ***Computing the extrinsic mean with the gradient descent algorithm.*** *In this example we computed the extrinsic mean μ_Φ of the withe dots on the manifold M^2 embedded in $\mathbb{R}^3$.*

Means on Manifolds (continues)

We prefer the more natural definition :

Definition (Intrinsic Mean)

Let M^p be a Riemannian manifold and $d(\cdot,\cdot)$ the geodesic distance. Then, the **intrinsic mean** of $x_1, \ldots, x_N \in M^p$ is :

$$\mu = \underset{x \in M^p}{\operatorname{argmin}} \sum_{i=1}^{N} d(x, x_i)^2.$$

- Existence and unicity if data well-localized.
- Can be computed by the gradient descent algorithm : $\nabla f(x) = -\frac{1}{N}\sum_{i=1}^{N} \operatorname{Log}_x(x_i)$,

  Update equation :

$$\mu_{j+1} = \operatorname{Exp}_{\mu_j}\left(\frac{\tau}{N}\sum_{i=1}^{N}\operatorname{Log}_{\mu_j}(x_i)\right),$$

- No optimal step size τ :

$$\nabla f(\operatorname{Exp}_{\mu_j}(\tau_j v)) = \nabla f(\mu_{j+1}) = 0.$$

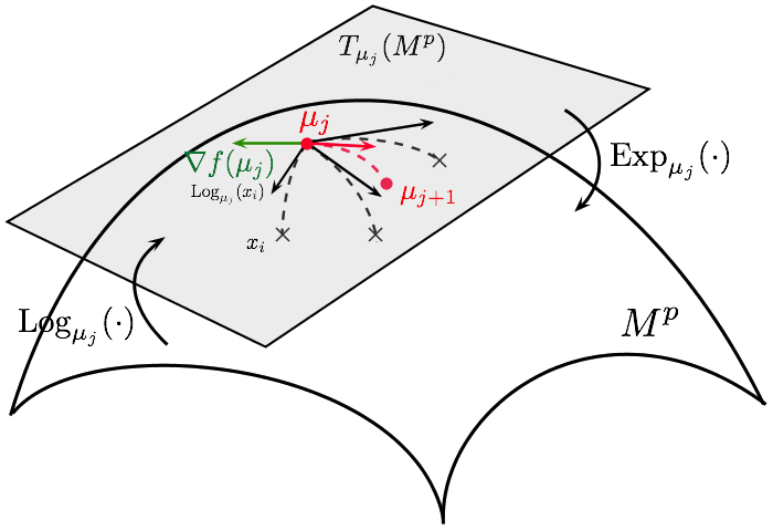


Figure: *One step of the gradient descent algorithm in the computation of the intrinsic mean.*

Variance, Geodesics submanifolds and Projection operator

Variance :
For $x \in \mathbb{R}^p$ one can show that :

$$\text{trace}\left(\text{Var}(x)\right) = \mathbb{E}\left[x^T x\right] - \mathbb{E}[x]^T \mathbb{E}[x].$$

which may be rewritten as :

$$\text{trace}\left(\text{Var}(x)\right) = \mathbb{E}\left[d(x,\mu)^2\right].$$

This inspires the natural definition :

Definition (Variance)

Let $d(\cdot,\cdot)$ the geodesic distance on M^p. Then, the **variance** of a r.v. $x \in M^p$ with mean μ is :

$$\sigma^2 = \mathbb{E}\left[d(x,\mu)^2\right].$$

Sample variance of $x_1, \ldots, x_N \in M^p$:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} d(\mu, x_i)^2 = \frac{1}{N}\sum_{i=1}^{N} \|\text{Log}_\mu(x_i)\|^2.$$

Geodesic submanifolds :

Definition (Geodesic Submanifolds)

A submanifold H of M^p is said to be **geodesic** at $x \in H$ if all geodesics of H passing through x are also geodesics of M^p.

A geodesic submanifold preserves the Riemannian distance : crucial for PGA !

Projection operator :

Definition (Projection operator)

Let $H \subset M^p$ a geodesic submanifold of M^p. Then, the **projection operator** $\pi_H : M^p \to H$ is :

$$\forall x \in M^p, \quad \pi_H(x) = \underset{y \in H}{\text{argmin}}\, d(x,y)^2,$$

Approximation formula (tangent plane) :

$$\text{Log}_p\left(\pi_H(x)\right) \simeq \sum_{i=1}^{k} \langle \text{Log}_p(x), v_i \rangle.$$

Principal Geodesics Analysis

Goal

Given a set $x_1, \cdots, x_N \in M^p$, find :

- A sequence of nested geodesic submanifolds $H_1 \subsetneq H_2 \subsetneq \cdots \subsetneq H_p = M^p$: the **principal geodesic submanifolds (PGS)**.
- A sequence of one-dimensional geodesic submanifolds $V_1, \cdots, V_p \subset M^p$: the **principal geodesic components (PGC)**.

Construction :
Let $v_1 \in T_\mu(M^p)$ be s.t. :

$$v_1 = \underset{\|v\|=1}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^{N} \|\mathrm{Log}_\mu(\pi_H(x_i))\|^2,$$

with $H = \mathrm{Exp}_\mu(\mathrm{span}\{v\} \cap U)$. The variance of the projected data is maximized.
We define :

- $H_1 = \mathrm{Exp}_\mu(\mathrm{span}\{v_1\} \cap U)$,
- $V_1 = \mathrm{Exp}_\mu(\mathrm{span}\{v_1\} \cap U)$.

Then, choose $v_2 \in T_\mu(M^p)$ s.t. :

$$v_2 = \underset{\|v\|=1}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^{N} \|\mathrm{Log}_\mu(\pi_H(x_i))\|^2,$$

with $H = \mathrm{Exp}_\mu(\mathrm{span}\{v_1, v\} \cap U)$. We define :

- $H_2 = \mathrm{Exp}_\mu(\mathrm{span}\{v_1, v_2\} \cap U)$,
- $V_2 = \mathrm{Exp}_\mu(\mathrm{span}\{v_2\} \cap U)$.

$\rightarrow$ Problem : the choice of v_2 isn't unique !

Principal geodesics Analysis (continues)

For unicity, we impose $v_2 \in \text{span}\{v_1\}^\perp$. This is not an arbitrary choice : attempt to de-correlate the data (analogously with PCA).
Weak de-correlation for two r.v. in $\mathbb{R}$: $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$. Into manifold setting :

$$\frac{1}{N}\sum_{i=1}^{N} \|\text{Log}_\mu(\pi_{H_2}(x_i))\|^2 = \frac{1}{N}\sum_{i=1}^{N} \|\text{Log}_\mu(\pi_{V_1}(x_i))\|^2 + \frac{1}{N}\sum_{i=1}^{N} \|\text{Log}_\mu(\pi_{V_2}(x_i))\|^2. \quad (4)$$

With $v_2 \in \text{span}\{v_1\}^\perp$ we have :

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} \|\text{Log}_\mu(\pi_{H_2}(x_i))\|^2 &\simeq \frac{1}{N}\sum_{i=1}^{N}(\langle v_1, \text{Log}_\mu(x_i)\rangle^2 + \langle v_2, \text{Log}_\mu(x_i)\rangle^2) \\ &\simeq \frac{1}{N}\sum_{i=1}^{N} \|\text{Log}_\mu(\pi_{V_1}(x_i))\|^2 + \frac{1}{N}\sum_{i=1}^{N} \|\text{Log}_\mu(\pi_{V_2}(x_i))\|^2.\end{aligned}$$

$\rightarrow$ We almost fulfilled (4) !

PGA : Interpretation

Interpretation

- Best n-dimensional geodesic submanifold to describe the data : the variance of the projected data is maximized.
- Attempt to re-express the data as a sequence of "independent" components (PCs analogs).

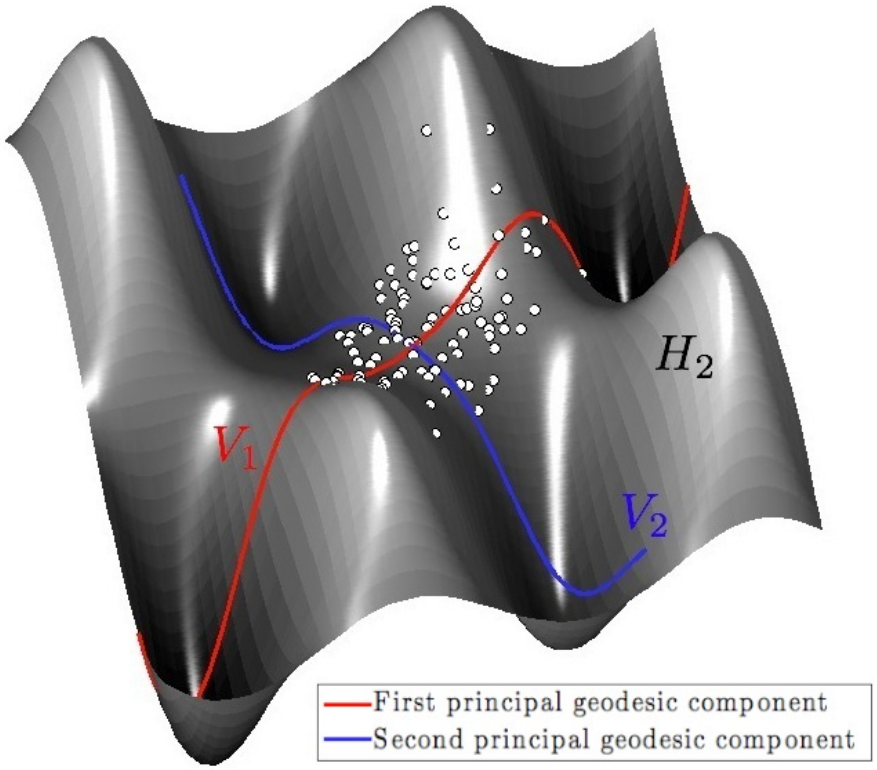


Figure: ***Example of a principal geodesic analysis.*** *The first principal geodesic submanifold is $H_1 = V_1$, the second H_2 is the manifold itself. The principal geodesic components V_1, V_2 are analog of the principal components in PCA.*

PGA : definition

Definition (Principal geodesic submanifolds/components)

Let M^p be a Riemannian manifold, $x_1, \cdots, x_N \in M^p$ and μ be the intrinsic mean of those points. Then, we define the p **principal geodesic submanifolds / components** as :

- **k-th principal geodesic submanifold / component :**
  Let $v_k \in T_\mu(M^p)$ be such that :

$$v_k = \underset{\substack{\|v\|=1 \\ v \in \mathrm{span}\{v_1, \cdots, v_{k-1}\}^\perp}}{\mathrm{argmax}} \frac{1}{N} \sum_{i=1}^{N} \|\mathrm{Log}_\mu(\pi_H(x_i))\|^2,$$

  with $H = \mathrm{Exp}_\mu(\mathrm{span}\{v_1, \cdots, v_{k-1}, v\} \cap U)$.
  Then, the **k-th principal geodesic submanifold** is :

$$H_k = \mathrm{Exp}_\mu(\mathrm{span}\{v_1, \cdots, v_{k-1}, v_k\} \cap U),$$

  and the **k-th principal geodesic component** is :

$$V_k = \mathrm{Exp}_\mu(\mathrm{span}\{v_k\} \cap U).$$

The shape space of triads

Let consider a triad $(x_1, x_2, x_3) \in \mathbb{C}^3$ such that $x_1 \neq x_2$. Its associated shape is the orbit of the triad under the symmetry group of *translations, rotations and dilatations* :

$$\mathcal{O}(x_1, x_2, x_3) = \{(wx_1 + z, wx_2 + z, wx_3 + z) \ : \ z \in \mathbb{C}, \ w \in \mathbb{C}^*\}.$$

We define the shape space of triads as : $\Sigma_2^3 = \{\mathcal{O}(x_1, x_2, x_3) \subset \mathbb{C}^3 : (x_1, x_2, x_3) \in \mathbb{C}^3\}$.

Parametrization :

Each triad is mapped to a point $z \in \mathbb{C}$:

$$z = \frac{2x_3 - (x_1 + x_2)}{x_2 - x_1}.$$

⚠ : Degenerate representation ! We need to add $z = \infty$ to complete it.

Spherical coordinates :

The stereographic projection (see figure), is a 1-1 correspondence between $\mathbb{C} \cup \{\infty\}$ and $S^2(1/2)$. A shape of a triad is then mapped to a point on the sphere :

$$\xi(z) = \left(\frac{1}{2}, \theta, \phi\right), \quad \theta \in [0, 2\pi] \ , \phi \in [0, \pi].$$

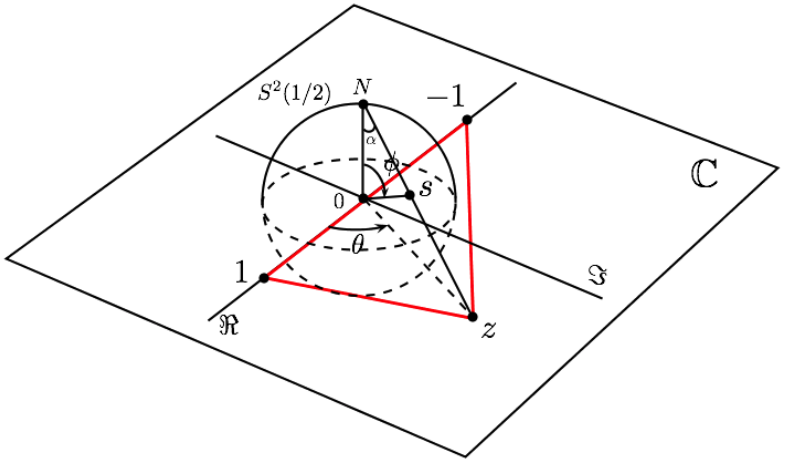


Figure: ***The stereographic projection.*** *The stereographic projection is a 1-1 correspondence between $\mathbb{C} \cup \{\infty\}$ and the sphere $S^2(1/2)$.*

The shape space of triads : representation

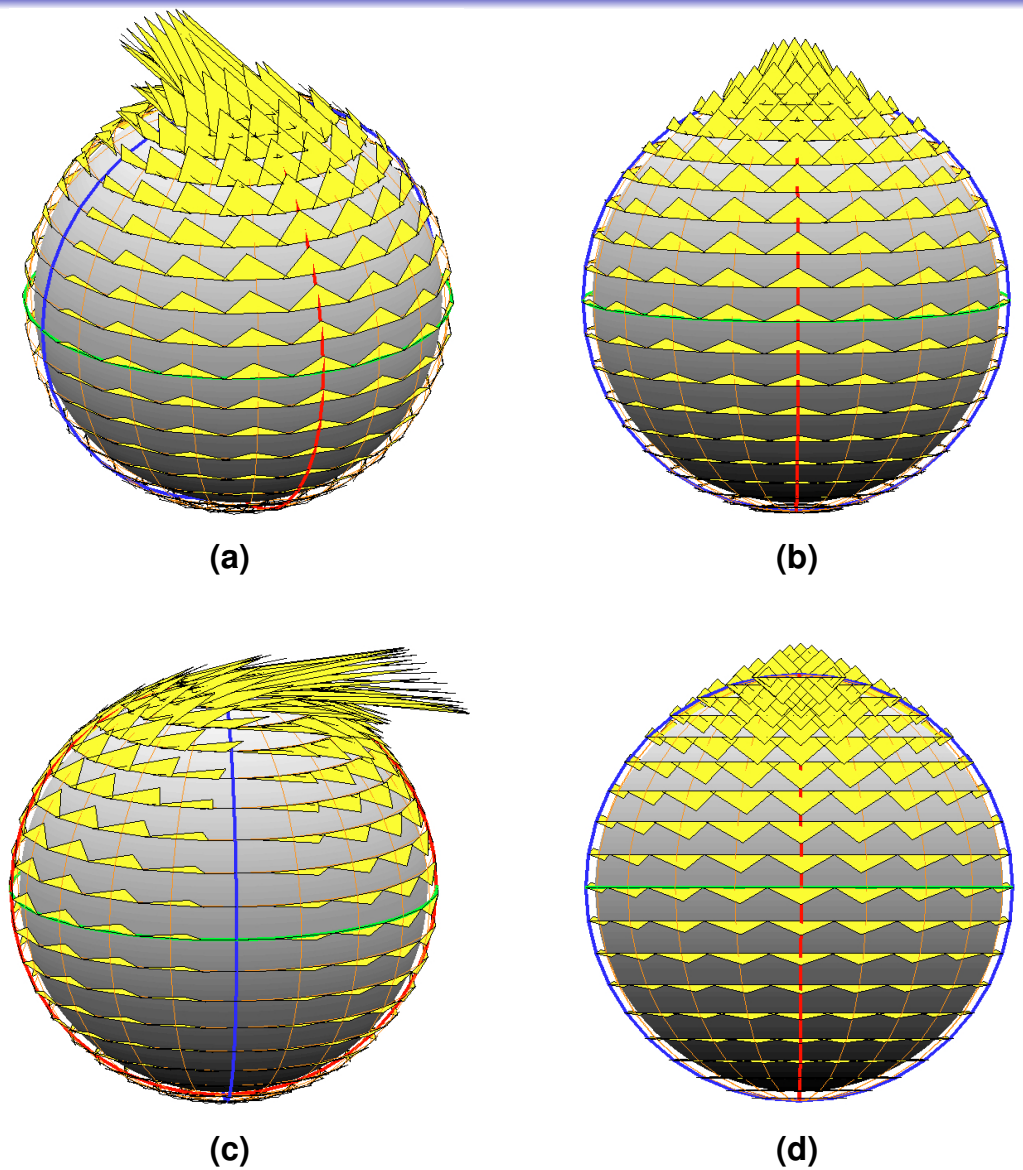


Figure: ***Different views of the shape space of planar triads.*** *The blue great circle corresponds to the great circle of collinear triads, while the red one corresponds to one of the great circles of isosceles triads. The green line is the equatorial line.*

PGA on the shape space of triads

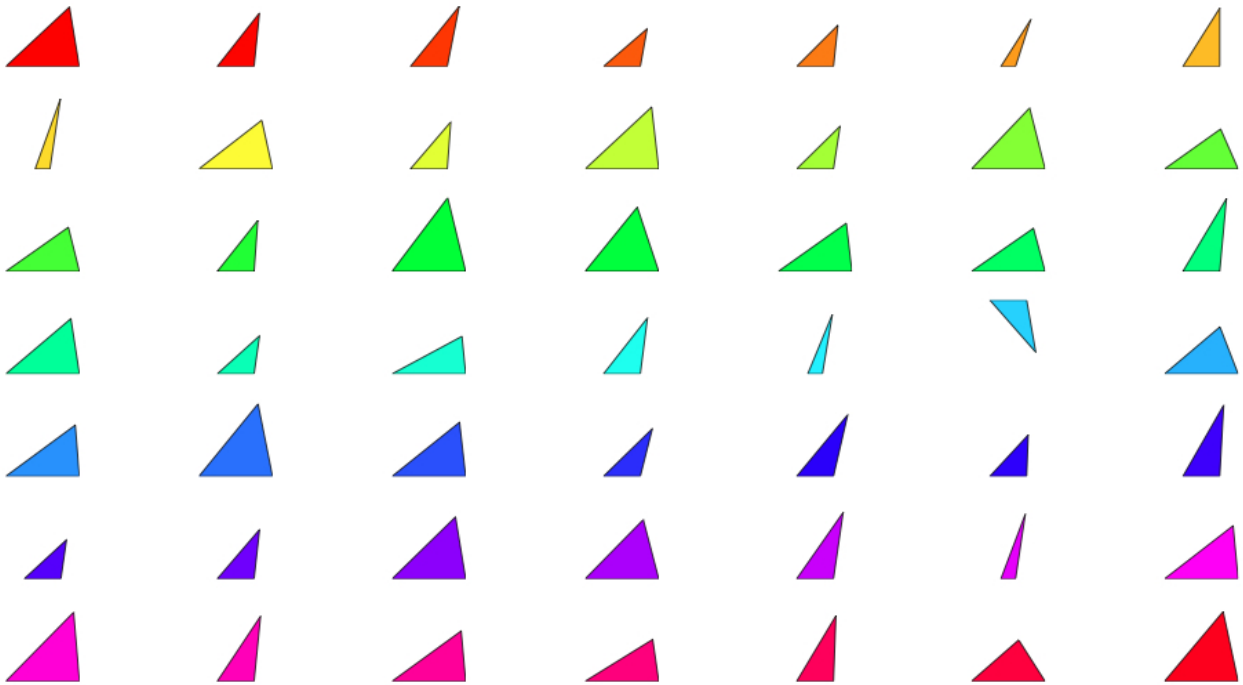


Figure: *A sample of 49 shapes randomly chosen in* Σ_2^3.

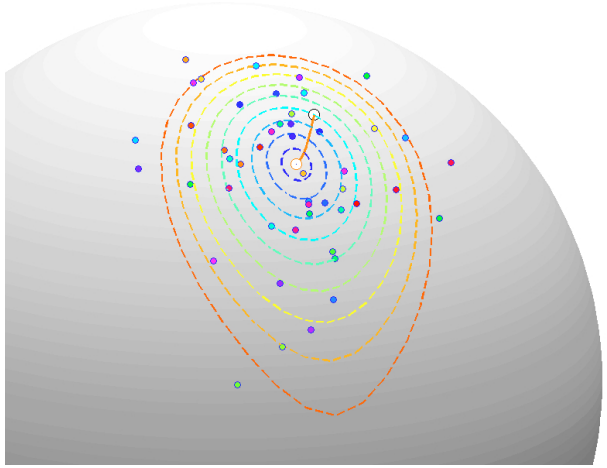
We wish to compute a principal geodesic analysis to efficiently describe the variability of the above sample of shapes.

Results of the PGA

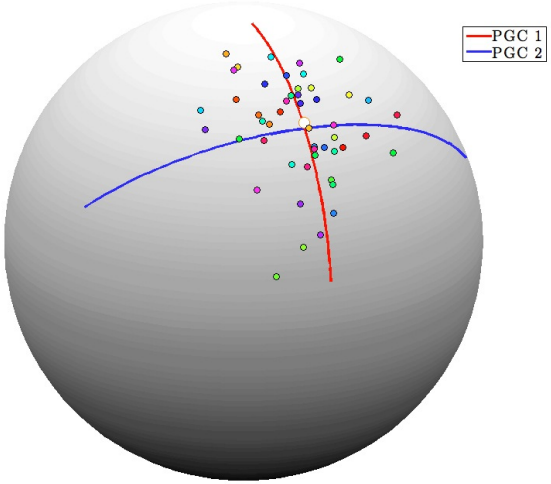
Numerical results

- Approximation of the intrinsic mean : $\nabla f(\mu_\phi) = (0.03, 0.01)$, thus $\mu_\phi \simeq \mu$.
- Variability along V_1 : 63.2%.

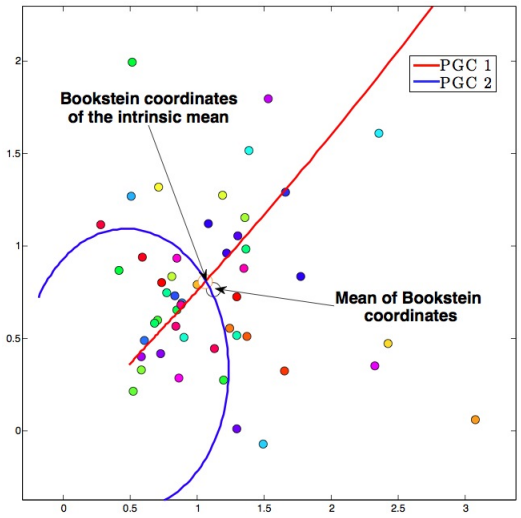
$\rightarrow$The analysis of variability is much more easier along the first geodesic component V_1.



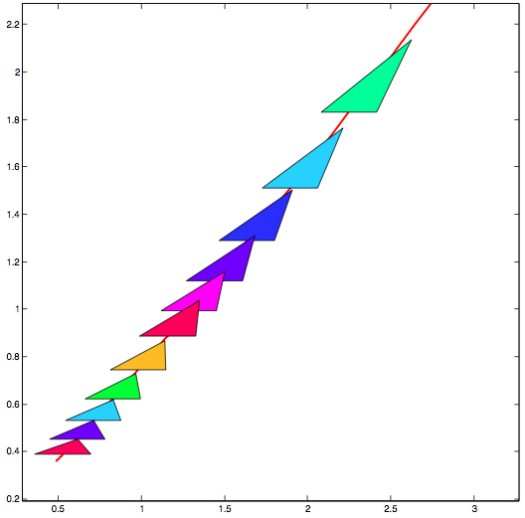
(a) *Computing extrinsic mean.*



(b) *PGCs V_1 and V_2.*



(c) *PGCs in $\mathbb{C} \cup \{\infty\}$*



(d) *Variation along V_1.*

Figure: ***Principal geodesic analysis of 49 shapes randomly chosen in Σ_2^3.***

Conclusion

- We successfully extended principal components analysis into manifold setting.
- We additionally tried to provide a generalization of principal components through principal geodesics components, allowing us to re-express the data as a sequence of "independent" components.
- However, our construction is based on an approximation.

Future work :

- Look for a construction independent from such an approximation.
- Pursue the construction of the shape space of triads Σ_2^3 in greater generality.
- Extend new statistical tools into manifold setting.

Bibliography

- C. G. Small *The Statistical Theory of Shape*, Springer series in statistics, 1996.
- Dietrich Stoyan, Wilfrid S. Kendall, Joseph Mecke, *Stochastic geometry and its applications*, Wiley series in probability and mathematical statistics, Second edition, 1995.
- I.L. Dryden and K.V. Mardia, *Statistical shape analysis*, Chichester : John Wiley & Sons, 2002.
- I. T. Joliffe, *Principal Component Analysis*, Second edition, Springer series in statistics, 2002.
- P. T. Fletcher, C. Lu, S. M. Pizer, S. Joshi, *Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape*, MIDAG.
- H.Karcher, *Riemannian center of mass and mollifier smoothing*, Communications on pure and applied mathematics, vol.30, no. 5, 1977.
- S. R. Buss and J. P. Fillmore, *Spherical averages and applications to spherical splines and interpolations*, ACM transactions on Graphics, vol 20, 2001.