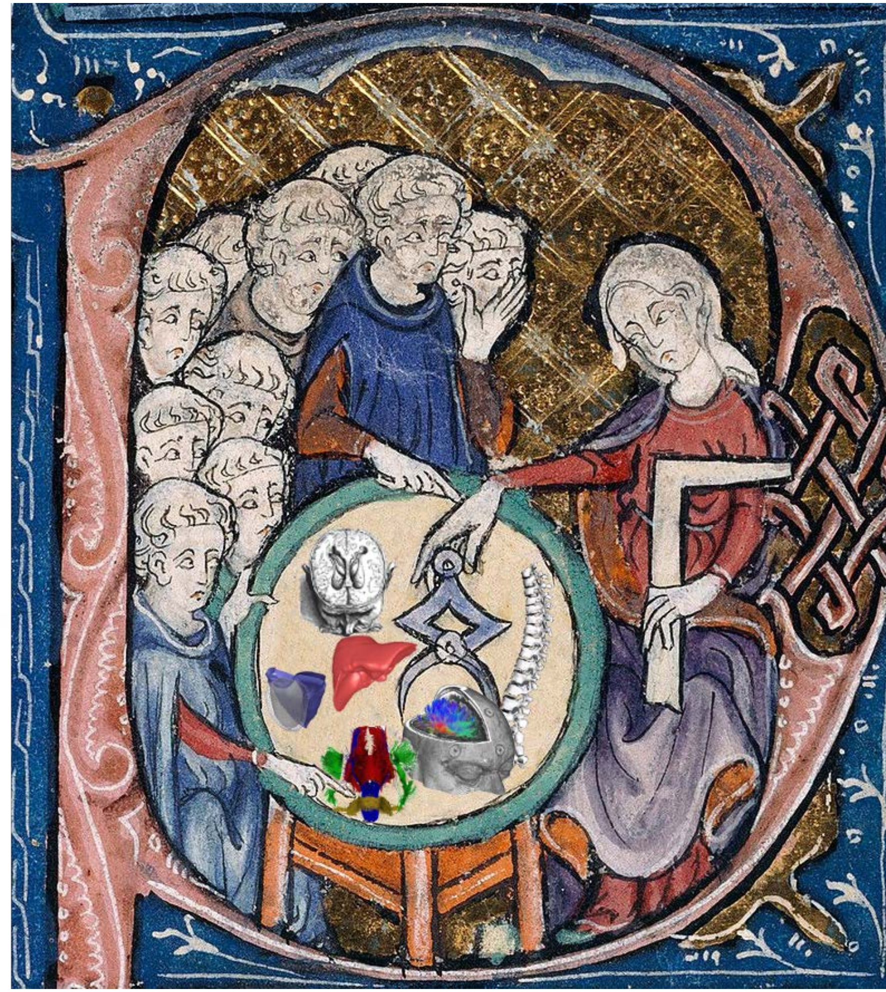
Xavier Pennec

Univ. Côte d'Azur and Inria, France



Freely adapted from "Women teaching geometry", in Adelard of Bath translation of Euclid's elements, 1310.

Geometric Statistics

Mathematical foundations and applications in computational anatomy

3/ Advanced Stats: empirical estimation and generalized PCA

Geometric Statistics workshop 09/2019

Geometric Statistics: Mathematical foundations and applications in computational anatomy

Intrinsic Statistics on Riemannian Manifolds

Metric and Affine Geometric Settings for Lie Groups

Advances Statistics: CLT & PCA

- Estimation of the empirical Fréchet mean & CLT
- Principal component analysis in manifolds
- Natural subspaces in manifolds: barycentric subspaces
- Rephrasing PCA with flags of subspaces

Several definitions of the mean

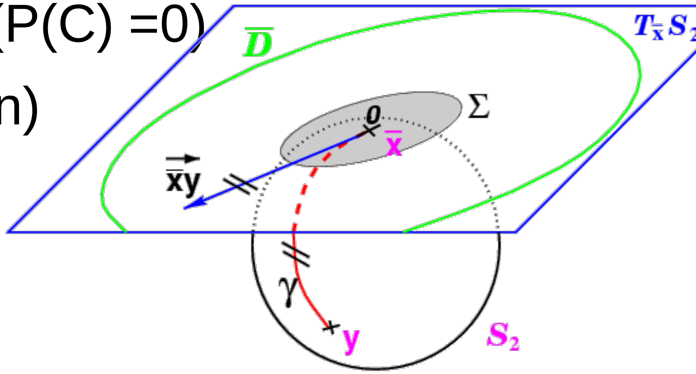
Tensor moments of a random point on M

- $\mathfrak{M}_1(x) = \int_M \overrightarrow{xz}\, dP(z)$ Tangent mean: (0,1) tensor field
- $\mathfrak{M}_2(x) = \int_M \overrightarrow{xz} \otimes \overrightarrow{xz}\, dP(z)$ 2nd moment: (0,2) tensor field
- $\mathfrak{M}_k(x) = \int_M \overrightarrow{xz} \otimes \overrightarrow{xz} \otimes \cdots \otimes \overrightarrow{xz}\, dP(z)$ k-contravariant tensor field
- $\sigma^2(x) = Tr_g\big(\mathfrak{M}_2(x)\big) = \int_M dist^2(x,z)\, dP(z)$ **Mean quadratic deviation**

Mean value = optimum of the variance

- **Frechet mean** [1944] = (global) minima of p-deviation (includes median)
- **Karcher mean** [1977] = local minima
- **Exponential barycenters** = critical points (P(C) =0)

  $\mathfrak{M}_1(\bar{x}) = \int_M \overrightarrow{\bar{x}z}\, dP(z) = 0$ (implicit definition)



Covariance at the mean

- $\Sigma = \mathfrak{M}_2(\bar{x}) = \int_M \overrightarrow{\bar{x}z} \otimes \overrightarrow{\bar{x}z}\, dP(z)$

Algorithms to compute the mean

Karcher flow (gradient descent)

$$\bar{x}_{t+1} = \exp_{\bar{x}_t}(\epsilon_t \, v_t) \;\; with \;\; v_t = \mathrm{E}(\overrightarrow{\mathrm{y}\mathbf{x}}) = \frac{1}{\mathrm{n}} \sum_i \log_{\bar{x}_t}(x_i)$$

- Usual algorithm with $\epsilon_t = 1$ can diverge on SPD matrices [Bini & Iannazzo, Linear Algebra Appl., 438:4, 2013]
- Convergence for non-negative curvature (p-means) [Afsari, Tron and Vidal, SICON 2013]

Inductive / incremental weighted means

- $\bar{x}_{k+1} = \exp_{\bar{x}_k}\left(\frac{1}{k} \, v_k\right) \;\; with \;\; v_k = \log_{\bar{x}_k}(x_{k+1})$
- On negatively curved spaces [Sturm 2003], BHV centroid [Billera, Holmes, Vogtmann, 2001]
- On non-positive spaces [G. Cheng, J. Ho, H. Salehian, B. C. Vemuri 2016]

Stochastic algorithm

- [Bonnabel IEE TAC 58(9) 2013]
- [Arnaudon & Miclo, Stoch. Proc. & App. 124, 2014]

Asymptotic behavior of the mean

Uniqueness of p-means with convex support

[Karcher 77 / Buser & Karcher 1981 / Kendall 90 / Afsari 10 / Le 11]

- Non-positively curved metric spaces (Aleksandrov): OK [Gromov, Sturm]
- Positive curvature: [Karcher 77 & Kendall 89] concentration conditions: Support in a regular geodesic ball of radius $r < r^* = \frac{1}{2}\min(inj(M), \pi/\sqrt{\kappa})$

Bhattacharya-Patrangenaru CLT [BP 2005, B&B 2008]

- Under suitable concentration conditions [KKC], for IID n-samples:
  - $\bar{x}_n \rightarrow \bar{x}$ *(consistency of empirical mean)*
  - $\sqrt{n}\, log_{\bar{x}}(\bar{x}_n) \rightarrow N(0, 4\bar{H}^{-1}\,\Sigma\,\bar{H}^{-1})$ *if* $\bar{H} = \int_M Hess_{\bar{x}}\left(d^2(y, \bar{x})\right) \mu(dy)$ *invertible*
- Problems for larger supports [Huckemann & Eltzner, H. Le]

Behavior in high concentration conditions?

- Interpretation of the mean Hessian?
- What happens for a small sample size (non-asymptotic behavior)?
- Can we extend results to affine connection spaces?

Concentration assumptions

- Uniqueness of the mean, support of diameter < ε

Riemannian manifold: Karcher & Kendall Concentr. Cond.

- $\mathrm{Supp}(\mu) \subset B(x,r)$ with $\mathrm{r} < \frac{1}{2}\, inj(x)$
- $\sup_{x \in B(x,r)} \kappa(x) < \pi^2/(4r)^2$

Affine connection spaces: Arnaudon & Li convexity cond.

- $\rho: M \times M \rightarrow R^+$ separating function
  - Separability: $\rho(x,y) = 0 \Leftrightarrow x = y$
  - Convexity along geodesic: $\rho(\gamma_1(t), \gamma_2(t)): R \rightarrow R^+ \; convex$
- p-convex geometry: $\mathrm{c\,dist^p(x,y)} \leq \rho(x,y) \leq C\, dist^p(x,y)$
- Uniqueness of exponential barycenter (compact support)

Taylor expansion in manifolds

The mean is an exponential barycenter

- The zero of the tangent mean field (Brewin Taylor expansion)

$$\mathfrak{M}_1(x) = \int_M \log_x(z)\, \mu(dz) \text{ has a zero at } \bar{x}.$$

Lots of additional terms in higher order derivatives since the vector field expression at $x_v = \exp_x(v)$ in a normal coordinates at x is modulated by $\mathrm{D}\exp_x(v)$.

- The zero of a mapping of vector spaces: the recenterd mean field

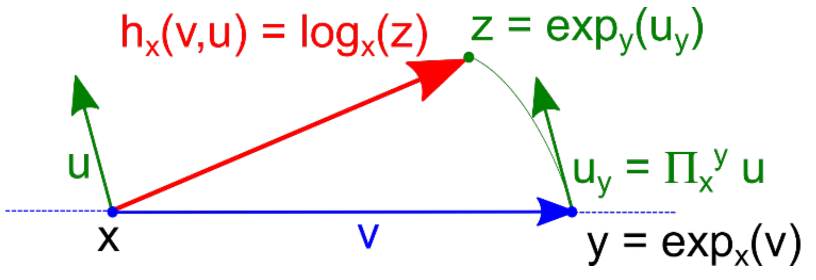
$$N_x(v) = \Pi_{x_v}^{x} \mathfrak{M}_1(\exp_x(v)) = \int_M \Pi_{x_v}^{x} \log_{x_v}(y)\, \mu(dy)$$

$\bar{x}$ is a Fréchet mean iff $N_x(v)$ has a zero at $\mathrm{v} = \log_x(\bar{x})$

Goal: compute a series expansion w.r.t. v

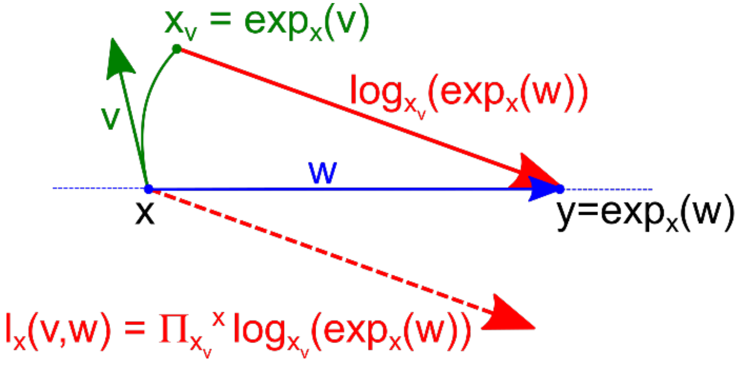
Taylor expansion in manifolds

Gavrilov's double exponential series (2006):



$$h_x(v,u) = \log_x(\Pi_x^{\exp_x(v)} u)$$
$$= v + u + \frac{1}{6} R(u,v)v + \frac{1}{3} R(u,v)u$$
$$+ \frac{1}{24} \nabla_v R(u,v)(2v + 5u)$$
$$+ \frac{1}{24} \nabla_u R(u,v)(v + 2u) + O(5)$$

Neighboring log expansion (new)



$$l_x(v,w) = \Pi_{x_v}^x \log_{x_v}(\exp_x(w))$$
$$= w - v + \frac{1}{6} R(w,v)(v - 2w)$$
$$+ \frac{1}{24} \nabla_v R(w,v)(2v - 3w)$$
$$+ \frac{1}{24} \nabla_w R(w,v)(v - 2w) + O(5)$$

Taylor expansion of recentered mean map

$$\mathfrak{N}_x^{\mu}(v) = \mathfrak{M}_1 - v + \frac{1}{6}R(\mathfrak{M}_1, v)v - \frac{1}{3}R(\bullet, v)\bullet : \mathfrak{M}_2 + \frac{1}{12}(\nabla_v R)(\mathfrak{M}_1, v)v$$
$$+ \frac{1}{24}(\nabla_\bullet R)(\bullet, v)v : \mathfrak{M}_2 - \frac{1}{8}(\nabla_v R)(\bullet, v)\bullet : \mathfrak{M}_2 - \frac{1}{12}(\nabla_\bullet R)(\bullet, v)\bullet : \mathfrak{M}_3 + O(\varepsilon^5)$$

Solving for the value $v = \log_x(\bar{x})$ that zeros the polynomial

$$\log_x(\bar{x}) = \mathfrak{M}_1 - \frac{1}{3}R(\bullet, \mathfrak{M}_1)\bullet : \mathfrak{M}_2 - \frac{1}{24}\nabla_\bullet R(\bullet, \mathfrak{M}_1)\mathfrak{M}_1 : \mathfrak{M}_2$$
$$- \frac{1}{8}\nabla_{\mathfrak{M}_1} R(\bullet, \mathfrak{M}_1)\bullet : \mathfrak{M}_2 - \frac{1}{12}\nabla_\bullet R(\bullet, \mathfrak{M}_1)\bullet : \mathfrak{M}_3 + O(\epsilon^5).$$

For an empirical an n-sample $\mathbf{X_n} = \frac{1}{n}\sum_i \delta_{x_i}$

$$\log_x(\bar{x}_n) = \mathfrak{X}_1^n - \frac{1}{3}R(\bullet, \mathfrak{X}_1^n)\bullet : \mathfrak{X}_2^n + \frac{1}{24}\nabla_\bullet R(\bullet, \mathfrak{X}_1^n)\mathfrak{X}_1^n : \mathfrak{X}_2^n$$
$$- \frac{1}{8}\nabla_{\mathfrak{X}_1^n} R(\bullet, \mathfrak{X}_1^n)\bullet : \mathfrak{X}_2^n - \frac{1}{12}\nabla_\bullet R(\bullet, \mathfrak{X}_1^n)\bullet : \mathfrak{X}_3^n + O(\epsilon^5).$$

Compute the expectation for a random n-sample?

Non-Asymptotic behavior of empirical means

Expectation of product of empirical moments

- $\mathbf{E}[\mathfrak{X}_k^n(x)] = \mathfrak{M}_k(x)$
- $\mathbf{E}\left[\mathfrak{X}_p^n \otimes \mathfrak{X}_q^n\right] = \frac{n-1}{n} \mathfrak{M}_{p+q} \otimes \mathfrak{M}_{p+q} + \frac{1}{n} \mathfrak{M}_{p+q}$
- Etc...

First Moment of the empirical mean

$$\begin{aligned}
\mathrm{E}\left[\, \log_x(\bar{x}_n) \,\right] &= \mathfrak{M}_1 - \frac{n-1}{3n} R(\bullet, \mathfrak{M}_1)\bullet \, \vdots \, \mathfrak{M}_2 \\
&+ \frac{(n-1)(n-2)}{24n^2} \left(\nabla_\bullet R(\bullet, \mathfrak{M}_1)\mathfrak{M}_1 \, \vdots \, \mathfrak{M}_2 - 3\nabla_{\mathfrak{M}_1} R(\bullet, \mathfrak{M}_1)\bullet \, \vdots \, \mathfrak{M}_2 \right) \\
&+ \frac{(n-1)}{12n^2} \left(2\nabla_\circ R(\circ, \bullet)\bullet \, \vdots \, \mathfrak{M}_2 \, \mathring{\vdots} \, \mathfrak{M}_2 - (1+n)\nabla_\bullet R(\bullet, \mathfrak{M}_1)\bullet \, \vdots \, \mathfrak{M}_3 \right) + O(\epsilon^5).
\end{aligned}$$

At the population mean:

$$\mathrm{E}\left[\, \log_{\bar{x}}(\bar{x}_n) \,\right] = \frac{(n-1)}{6n^2} \nabla_\bullet R(\bullet, \circ)\circ \, \vdots \, \mathfrak{M}_2 \, \mathring{\vdots} \, \mathfrak{M}_2 + O(\epsilon^5).$$

Non-Asymptotic behavior of empirical means

Second Moment of the empirical mean a the pop. mean:

$$\mathbf{E}\left[\,\log_{\bar{x}}(\bar{x}_n)\otimes\log_{\bar{x}}(\bar{x}_n)\,\right]=\frac{1}{n}\mathfrak{M}_2-\frac{(n-1)}{3n^2}\mathfrak{M}_2 \mathbin{\overset{\circ}{\circ}} \left(\circ\otimes R(\bullet,\circ)\bullet+R(\bullet,\circ)\bullet\otimes\circ\right) \mathbin{\overset{\bullet}{\bullet}} \mathfrak{M}_2+O(\epsilon^5).$$

In coordinates:

$$\begin{aligned} Bias(\bar{x}_n)^a &= \tfrac{1}{6n}\left(1-\tfrac{1}{n}\right)\nabla_b R^a_{cde}\mathfrak{M}_2^{ce}\mathfrak{M}_2^{bd}+O\left(\epsilon^5\right) \\ Cov(\bar{x}_n)^{ab} &= \tfrac{1}{n}\left(\mathfrak{M}_2^{ab}-\tfrac{1}{3}\left(1-\tfrac{1}{n}\right)\mathfrak{M}_2^{cd}\left(\mathfrak{M}_2^{ae}R^b_{cde}+R^a_{cde}\mathfrak{M}_2^{be}\right)\right)+O\left(\epsilon^5\right). \end{aligned}$$

Non-Asymptotic behavior of empirical means

Moments of the Fréchet mean of a n-sample

- **Unexpected bias** in 1/n on empirical mean (**gradient of curvature-cov.**)

$$\mathbf{bias}(\bar{x}_n) = \boldsymbol{E}\left(log_{\bar{x}}(\bar{x}_n)\right) = \frac{\mathbf{1}}{\boldsymbol{6n}}\left(\mathfrak{M}_2 : \nabla \boldsymbol{R} : \mathfrak{M}_2\right) + O(\epsilon^5, 1/n^2)$$

- **Concentration rate** modulated by the **curvature-covariance**:

$$\boldsymbol{Cov}(\bar{x}_n) = \boldsymbol{E}\left(log_{\bar{x}}(\bar{x}_n) \otimes log_{\bar{x}}(\bar{x}_n)\right) = \frac{\mathbf{1}}{\boldsymbol{n}}\mathfrak{M}_2 + \frac{\mathbf{1}}{\boldsymbol{3n}}\mathfrak{M}_2 : \boldsymbol{R} : \mathfrak{M}_2 + O(\epsilon^5, 1/n^2)$$

  - **Asymptotically infinitely fast** CV for negative curvature
  - **No convergence (LLN fails)** at the limit of KKC condition

[XP, Curvature effects on the empirical mean in Manifolds 2019, arXiv:1906.07418]

Comparison with the BP CLT

Bhattacharya-Patrangenaru CLT [BP 2005, B&B 2008]

- Under suitable concentration conditions, for IID n-samples:
  - $\bar{x}_n \to \bar{x}$ *(consistency of empirical mean)*
  - $\sqrt{n}\, log_{\bar{x}}(\bar{x}_n) \to N(0, 4\bar{H}^{-1}\, \Sigma\, \bar{H}^{-1})$ *if* $\bar{H} = \int_M Hess_{\bar{x}}\big(\, d^2(y,\bar{x})\big)\, \mu(dy)$ *invertible*

Hessian: $\frac{1}{2}\bar{H} = Id + \frac{1}{3} R : \mathfrak{M}_2 + \frac{1}{12} \nabla \mathrm{R} : \mathfrak{M}_3 + O(\epsilon^4, 1/n^2)$

$$4[\bar{H}^{(-1)}\mathfrak{M}_2\bar{H}^{(-1)}]^{ab} = 4[\bar{H}^{(-1)}]^a_c[\mathfrak{M}_2]^{cd}[\bar{H}^{(-1)}]^b_d = \mathfrak{M}_2^{ab} - \frac{1}{3}\mathfrak{M}_2^{ef}\left(R^a_{efc}\mathfrak{M}_2^{cb} + \mathfrak{M}_2^{ad}R^b_{efd}\right) + O(\varepsilon^5),$$

Same limiting expansion for large n

Isotropic distribution in constant curvature spaces

- Symmetric spaces: no bias
- Variance is modulated w.r.t. Euclidean: $Var(\bar{x}_n) = \alpha \frac{\sigma^2}{n}$

High concentration expansion

- $\alpha = 1 + \frac{2}{3}\left(1 - \frac{1}{d}\right)\left(1 - \frac{1}{n}\right)\kappa\sigma^2 + O(\epsilon^5)$

Closed form for asymptotic BP-CLT expansion

$$\frac{1}{2} H_x(y) = uu^\top + h(\kappa\theta^2)(\mathrm{Id} - uu^\top) \qquad \text{with} \qquad h(t) = \sqrt{t}\cot(\sqrt{t})$$

- $\alpha = \left(\frac{1}{d} + \left(1 - \frac{1}{d}\right)\bar{h}\right)^{-2} + O(n^{-2})$

$$\bar{h} = \mathbf{E}\left[\, h(\kappa \operatorname{dist}(\bar{x}, .)^2)\,\right] = \int_{\mathcal{M}} h(\kappa(\operatorname{dist}(\bar{x}, y)^2)\, \mu(dy).$$

Isotropic distribution in constant curvature spaces

- Variance is modulated w.r.t. Euclidean: $Var(\bar{x}_n) = \alpha \frac{\sigma^2}{n}$

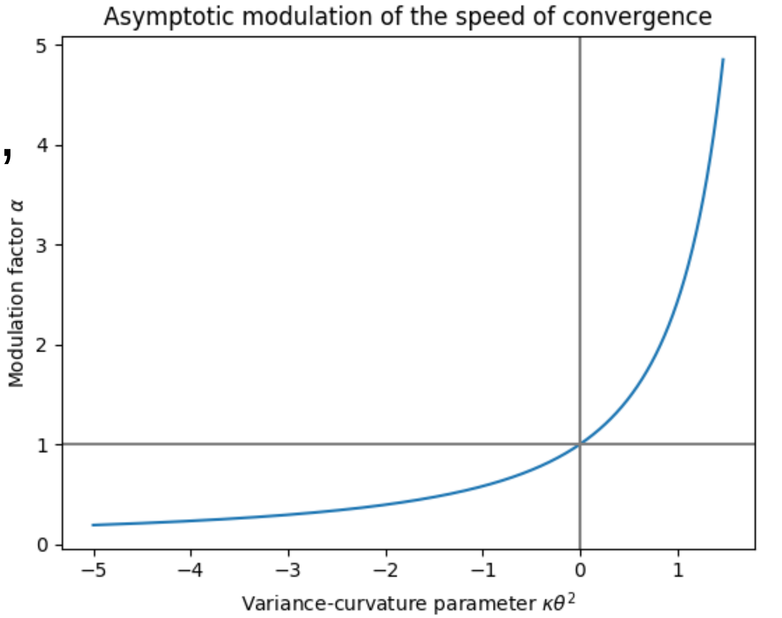
)

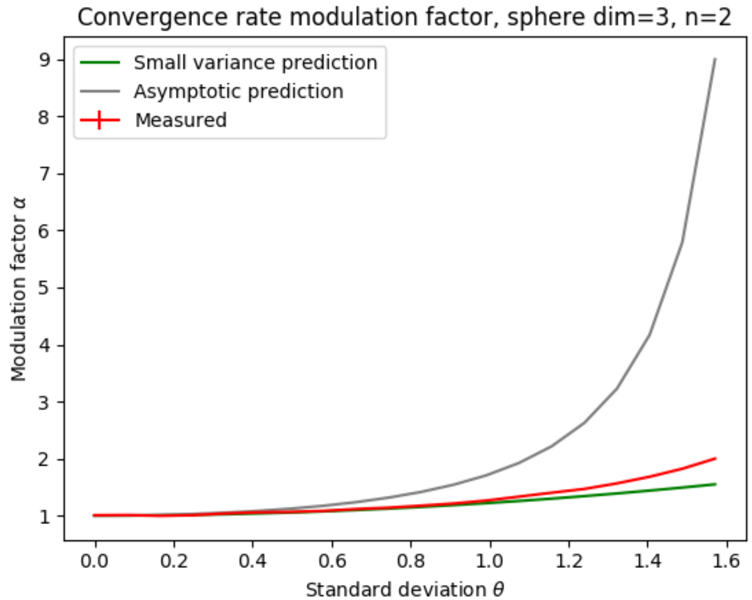
Asymptotic BP-CLT expansion

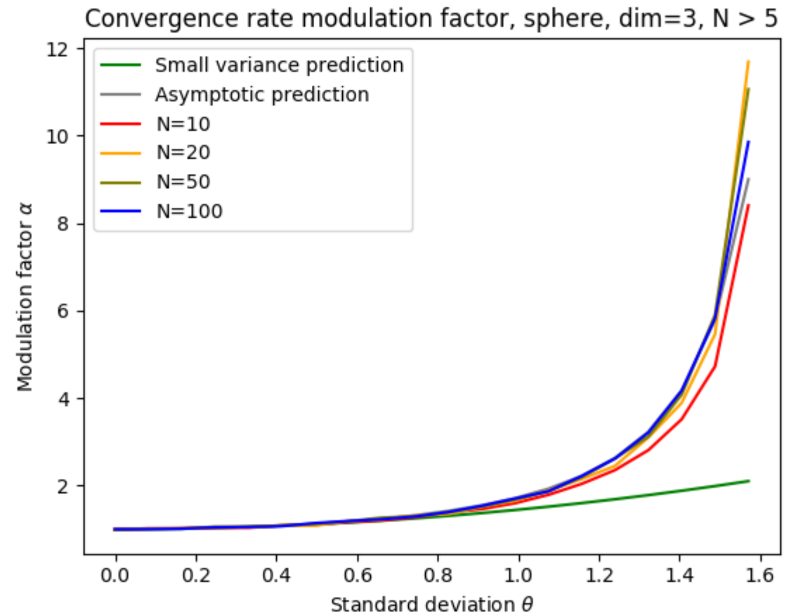
- $\alpha = \left(\frac{1}{d} + \left(1 - \frac{1}{d}\right)\bar{h}\right)^{-2} + O(n^{-2})$

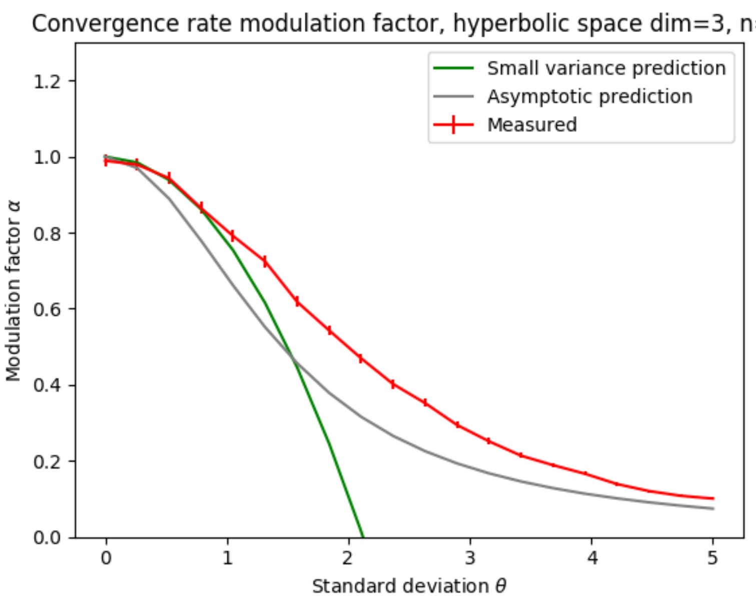
Archetypal modulation factor

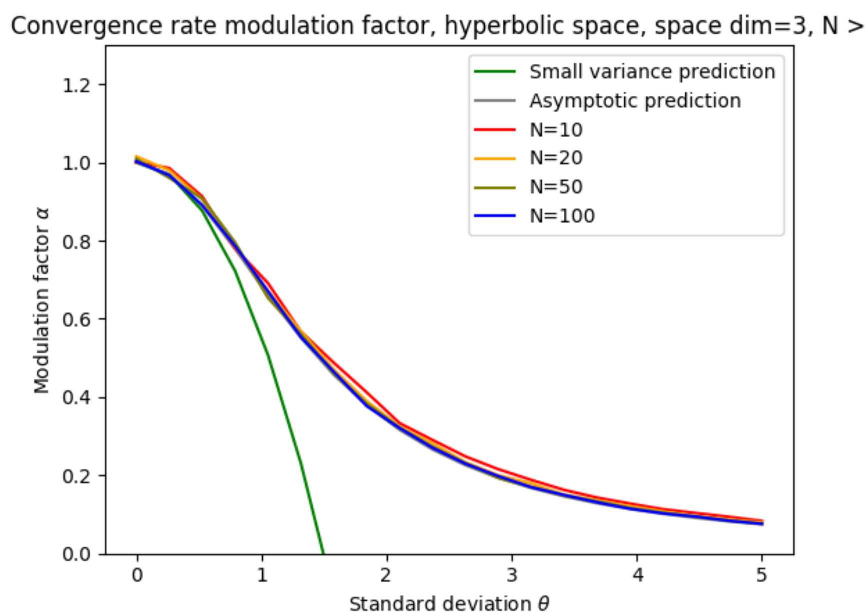
- Uniform distrib on $S(\bar{x}, \theta) \subset M$, large n, large d
- $\alpha = \frac{\tan^2\left(\sqrt{\kappa\theta^2}\right)}{\kappa\theta^2}$




Convergence rate modulation factor, sphere dim=3, n=2
Small variance prediction
Asymptotic prediction
Measured
Modulation factor α
Standard deviation θ
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1
2
3
4
5
6
7
8
9


Convergence rate modulation factor, sphere, dim=3, N > 5
Small variance prediction
Asymptotic prediction
N=10
N=20
N=50
N=100
Modulation factor α
Standard deviation θ
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
2
4
6
8
10
12


Convergence rate modulation factor, hyperbolic space dim=3, n=2
Small variance prediction
Asymptotic prediction
Measured
Modulation factor α
Standard deviation θ
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
1.2


Convergence rate modulation factor, hyperbolic space, space dim=3, N > 5
Small variance prediction
Asymptotic prediction
N=10
N=20
N=50
N=100
Modulation factor α
Standard deviation θ
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
1.2

Conclusions

High concertation expansion very accurate for low theta

Asymptotic expansion very accurate for n> 10

Main variable controlling the modulation is variance-curvature tensor

$$R(\blacksquare, {}^{\circ})\blacksquare : \mathfrak{M}_2$$

Main variable controling the bias

$$\mathfrak{M}_2 : \nabla \circ R({}^{\circ}, \blacksquare)\blacksquare : \mathfrak{M}_2$$

Geometric Statistics: Mathematical foundations and applications in computational anatomy

Intrinsic Statistics on Riemannian Manifolds

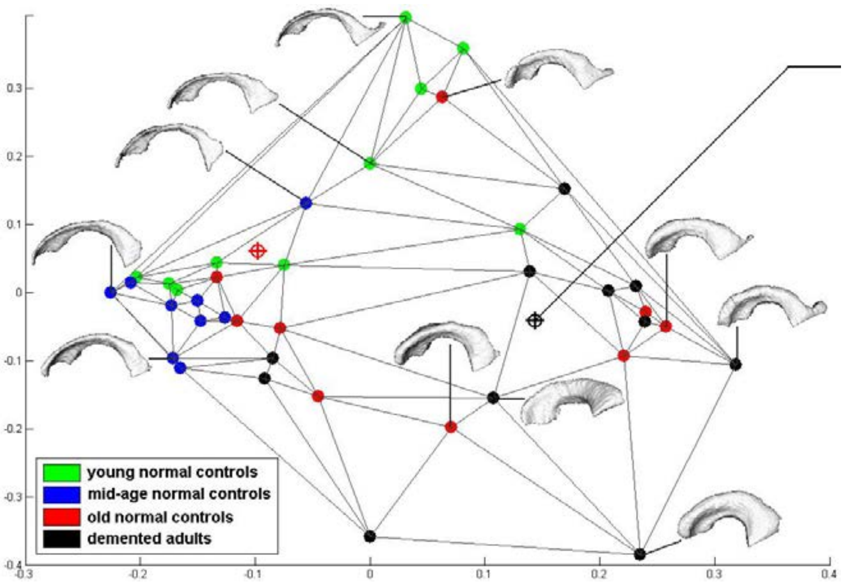
Metric and Affine Geometric Settings for Lie Groups

Parallel Transport to Analyze Longitudinal Deformations

Advances Statistics: CLT & PCA

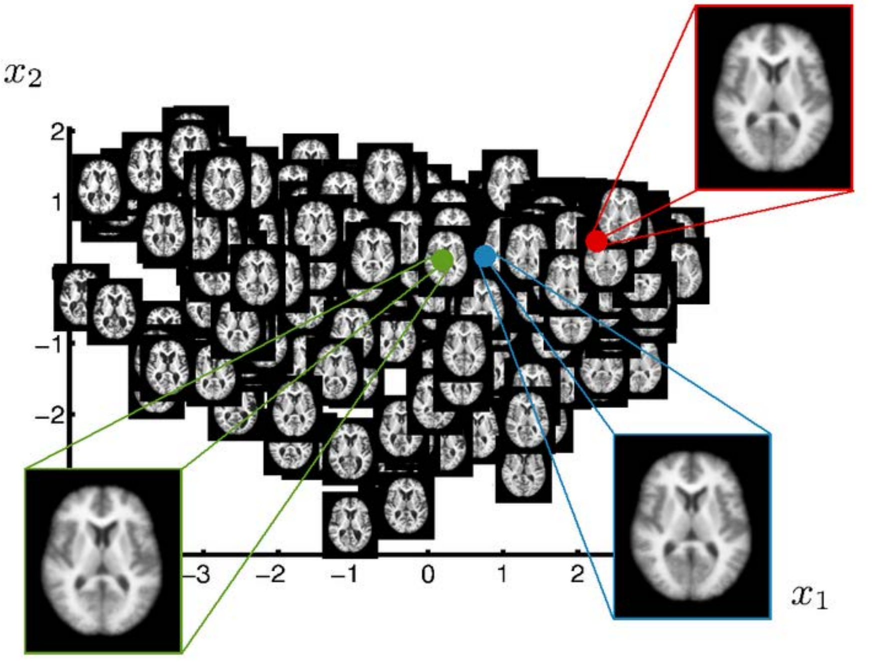
- Estimation of the empirical Fréchet mean & CLT
- Principal component analysis in manifolds
- Natural subspaces in manifolds: barycentric subspaces
- Rephrasing PCA with flags of subspaces

Low dimensional subspace approximation?



Manifold of cerebral ventricles

Etyngier, Keriven, Segonne 2007.



Manifold of brain images

S. Gerber et al, Medical Image analysis, 2009.

- Beyond the 0-dim mean → higher dimensional subspaces
- When embedding structure is already manifold (e.g. Riemannian): Not manifold learning (LLE, Isomap,…) but **submanifold learning**
- **Natural subspaces for extending PCA to manifolds?**

Tangent PCA (tPCA)

Maximize the squared distance to the mean (explained variance)

- Algorithm
  - Unfold data on tangent space at the mean
  - Diagonalize covariance at the mean $\Sigma(x) \propto \sum_i \overrightarrow{\bar{x}x_i}\,\overrightarrow{\bar{x}x_i}^t$

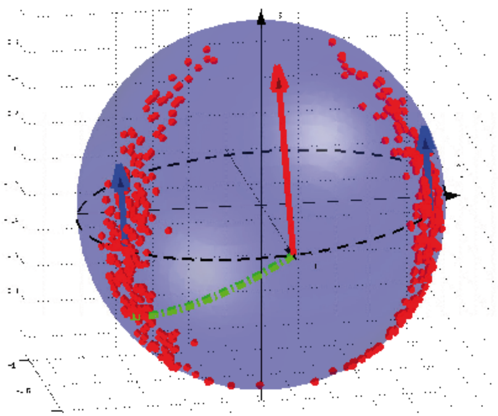
- Generative model:
  - Gaussian (large variance) in the horizontal subspace
  - Gaussian (small variance) in the vertical space

- Find the subspace of $T_x M$ that best explains the variance

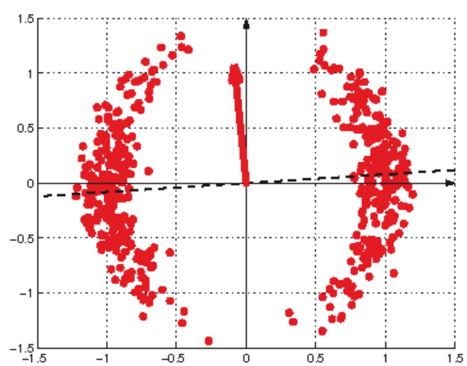
Problems of tPCA

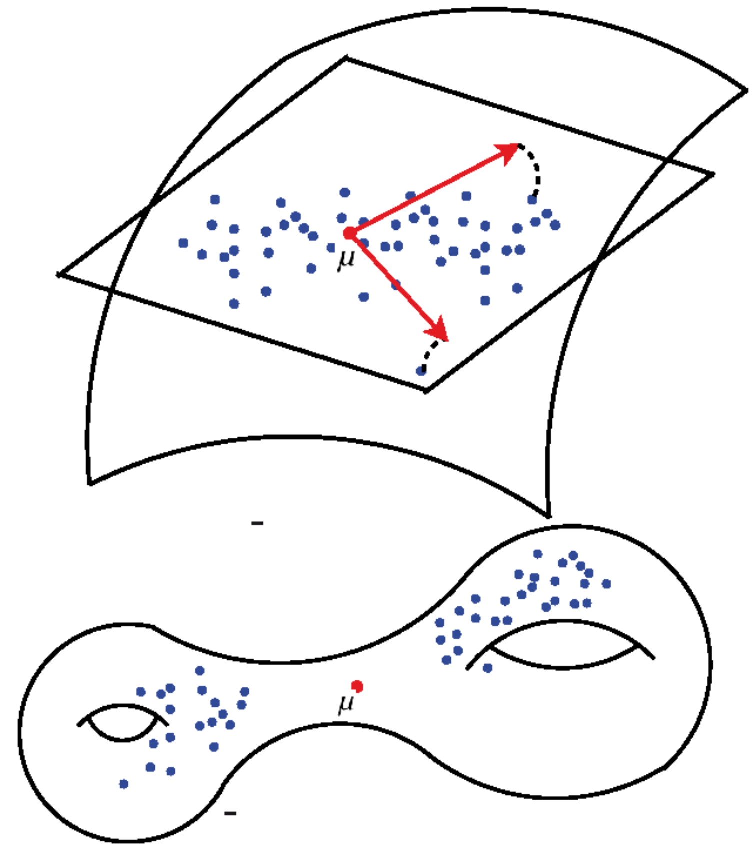
Analysis is done relative to the mean

- What if the mean is a poor description of the data?
  - Multimodal distributions
  - Uniform distribution on subspaces
  - Large variance w.r.t curvature



Bimodal distribution on S2





Images courtesy of S. Sommer

Principal Geodesic / Geodesic Principal Component Analysis

Minimize the squared Riemannian distance to a low dimensional subspace (unexplained variance)

- **Geodesic Subspace**: $GS(x, w_1, \dots w_k) = \{\exp_x(\sum_i \alpha_i w_i) \ for \ \alpha \in R^k \}$
  - Parametric subspace spanned by geodesic rays from point x
  - Beware: GS have to be restricted to be well posed [XP, AoS 2018]
    - PGA (Fletcher et al., 2004, Sommer 2014)
    - Geodesic PCA (GPCA, Huckeman et al., 2010)

- Generative model:
  - Unknown (uniform ?) distribution within the subspace
  - Gaussian distribution in the vertical space

Asymmetry w.r.t. the base point in $GS(x, w_1, \dots w_k)$

- Totally geodesic at x only

Patching the Problems of tPCA / PGA

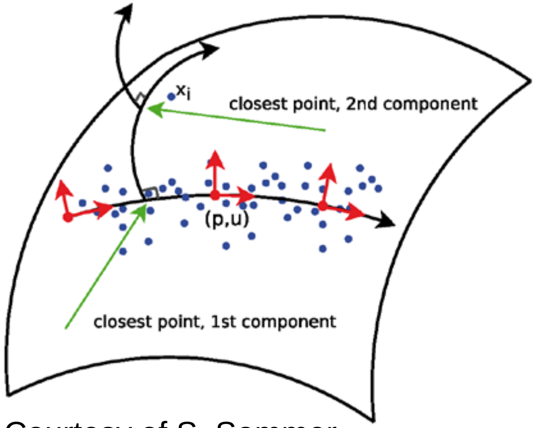
Improve the flexibity of the geodesics

- 1D regression with higher order splines **[Gu, Machado, Leite, Vialard, Singh, Niethammer, Absil,…]**
  - **Control of dimensionality for n-D Polynomials on manifolds?**

Iterated Frame Bundle Development

[HCA, Sommer GSI 2013]

- Iterated construction of subspaces
- Parallel transport in frame bundle
  - **Intrinsic asymmetry between components**



Courtesy of S. Sommer

Nested "algebraic" subspaces

- Principal nested spheres **[Jung, Dryden, Marron 2012]**
- Quotient of Lie group action **[Huckemann, Hotz, Munk, 2010]**
  - **No general semi-direct product space structure in general Riemannian manifolds**

Geometric Statistics: Mathematical foundations and applications in computational anatomy

Intrinsic Statistics on Riemannian Manifolds

Metric and Affine Geometric Settings for Lie Groups

Parallel Transport to Analyze Longitudinal Deformations

Advances Statistics: CLT & PCA

- Estimation of the empirical Fréchet mean & CLT
- Principal component analysis in manifolds
- Natural subspaces in manifolds: barycentric subspaces
- Rephrasing PCA with flags of subspaces

Affine span in Euclidean spaces

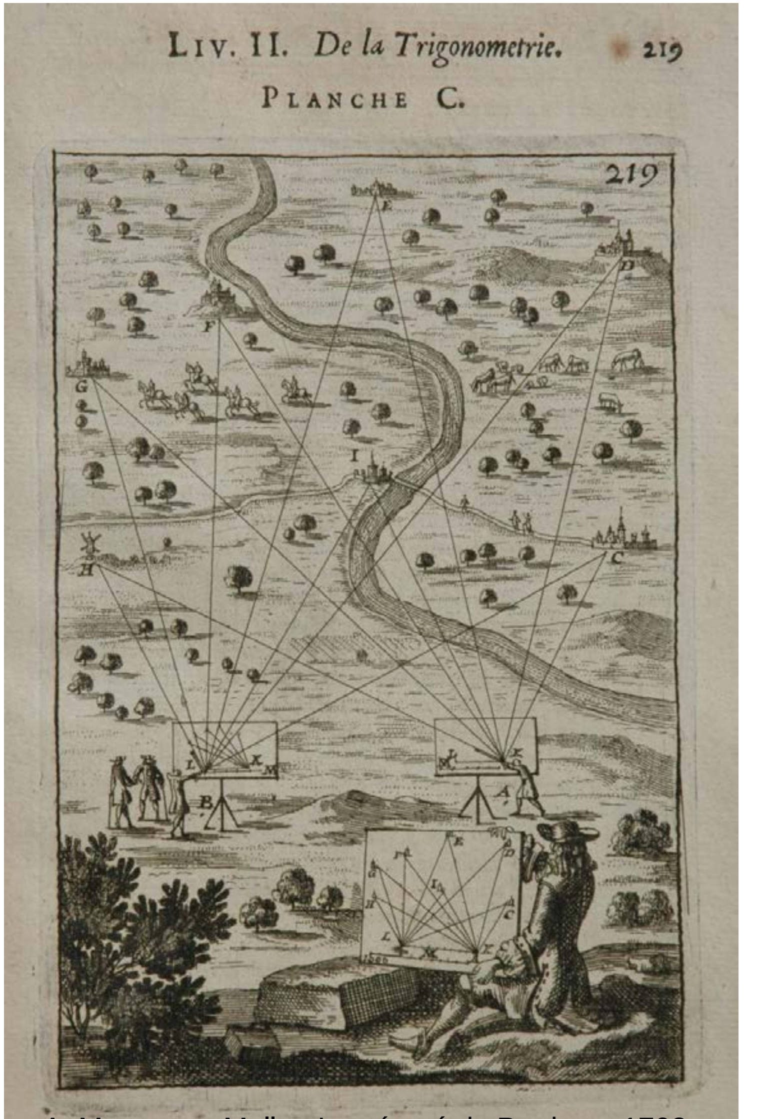
Affine span of (k+1) points: weighted barycentric equation

$$\mathrm{Aff}(x_0, x_1, \dots x_k) = \{x = \sum_i \lambda_i x_i \;\; with \;\sum_i \lambda_i = 1\}$$
$$= \{x \in R^n \;\; s.t \;\sum_i \lambda_i (x_i - x) = 0, \lambda \in P_k^*\}$$

Key ideas:

- ~~tPCA, PGA: Look at data points from the mean (mean has to be unique)~~

- Triangulate from several reference: **locus of weighted means**



A. Manesson-Mallet. La géométrie Pratique, 1702

Barycentric subspaces and Affine span in Riemannian manifolds

Fréchet / Karcher barycentric subspaces (KBS / FBS)

- Normalized weighted variance: $\sigma^2(x,\lambda) = \sum \lambda_i dist^2(x, x_i) / \sum \lambda_i$
- Set of absolute / local minima of the λ-variance
- Works in stratified spaces (may go accross different strata)
  - Non-negative weights: Locus of Fréchet Mean [Weyenberg, Nye]

Exponential barycentric subspace and affine span

- Weighted exponential barycenters: $\mathfrak{M}_1(x,\lambda) = \sum_i \lambda_i \overrightarrow{xx_i} = 0$
- EBS$(x_0, \dots x_k) = \{ x \in M^*(x_0, \dots x_k) \mid \mathfrak{M}_1(x,\lambda) = 0 \}$
- Affine span = closure of EBS in M $\quad Aff(x_0, \dots x_k) = \overline{EBS(x_0, \dots x_k)}$

Questions

- Local structure: local manifold? dimension? stratification?
- Relationship between KBS ⊂ FBS, EBS and affine span?

[X.P. Barycentric Subspace Analysis on Manifolds. Annals of statistics. 2018. arXiv:1607.02833]

Analysis of Barycentric Subspaces

Assumptions:

- Restrict to the **punctured manifold** $M^*(x_0, \dots x_k) = M \,/ \cup C(x_i)$
  - $dist^2(x, x_i)$, $\log_x(x_i)$ are smooth but M^* may be split in pieces
- Affinely independent points:
  $\{\overrightarrow{x_i x_j}\}_{0 \le i \neq j \le k}$ exist and are linearly independent for all i

Local well posedness for the barycentric simplex:

- EBS / KBS are well defined in a neighborhood of reference points
- For reference points in a sufficiently small ball and positive weights: unique Frechet = Karcher = Exp Barycenter in that ball: smooth graph of a k-dim function [proof using Buser & Karcher 81]

SVD characterization of EBS: $\mathfrak{M}_1(\mathrm{x}, \lambda) = Z(x)\lambda = 0$

- SVD: $Z(x) = [\,\overrightarrow{x x_0}, \dots \overrightarrow{x x_k}] = U(x)S(x)V^t(x)$
  - $EBS(x_0, \dots x_k) =$ Zero level-set of l>0 singular values of Z(x)
  - Stratification on the number of vanishing singular values

[X.P. Barycentric Subspace Analysis on Manifolds. Annals of statistics. 2018. arXiv:1607.02833]

Analysis of Barycentric Subspaces

Exp. barycenters are critical points of λ-variance on M*

- $\nabla\sigma^2(x,\lambda) = -2\mathfrak{M}_1(x,\lambda) = 0$ $\qquad$ $\boldsymbol{KBS \cap M^* \subset EBS}$

Caractérisation of local minima: Hessian (if non degenerate)

$$H(x,\lambda) = -2\sum_i \lambda_i D_x \log_x(x_i) = \mathbf{Id} - \frac{\mathbf{1}}{\mathbf{3}}\mathbf{Ric}\left(\mathfrak{M}_2(\mathbf{x},\boldsymbol{\lambda})\right) + \text{HOT}$$

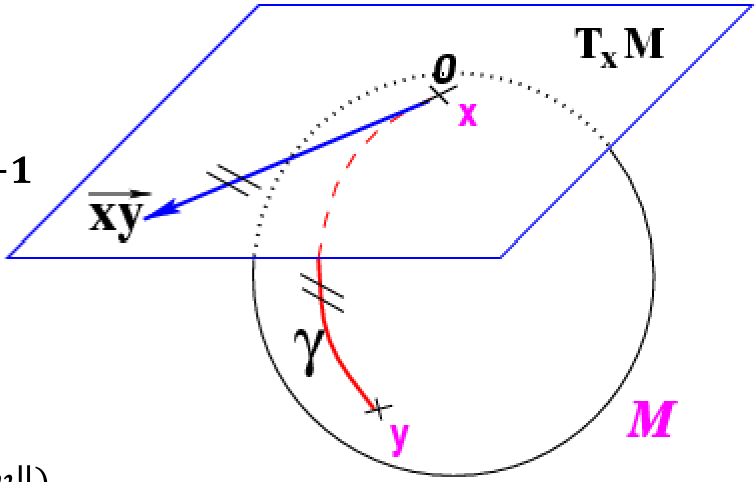
Regular and positive pts (non-degenerated critical points)

- $\boldsymbol{EBS^{Reg}(x_0, \ldots x_k) = \{\, x \in Aff(x_0, \ldots x_k), s.t.\ H(x, \lambda^*(x)) \neq 0\}}$
- $\boldsymbol{EBS^{+}(x_0, \ldots x_k) = \{\, x \in Aff(x_0, \ldots x_k), s.t.\ H(x, \lambda^*(x))\ Pos.def.\}}$

Theorem: EBS partitioned into cells by the index of the Hessian of λ-variance: KBS = EBS^+ on M*

[X.P. Barycentric Subspace Analysis on Manifolds. Annals of statistics. 2018. arXiv:1607.02833]

Example on the sphere



Manifold

- Unit sphere $\mathcal{M} = \boldsymbol{S_n}$ embedded in $\mathbb{R}^{n+1}$
- ||x|| =1

Exp and log map

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

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

Distance

$$dist(x, y) = \|\log_x(y)\| = \theta$$

(k+1)-pointed & punctured Sphere

- $X = [x_0, x_1, \dots, x_k] \in (S_n)^k$
- Punctured sphere: exclude antipodal points: $S_n^* = S_n / -X$

KBS / FBS with 3 points on the sphere

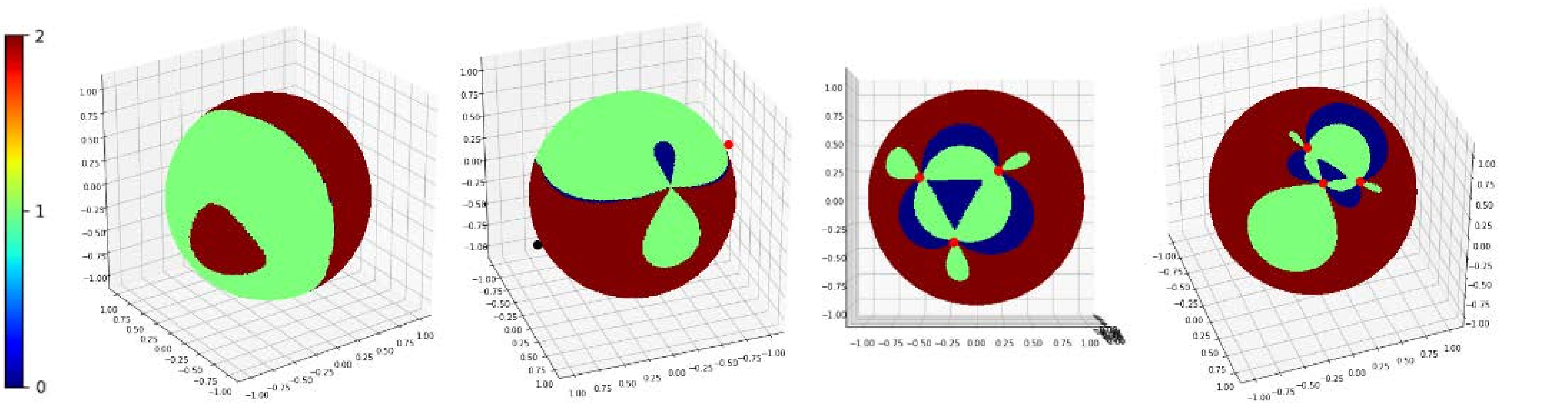
EBS: great subspheres spanned by reference points (mod cut loci)

$$\text{EBS}(x_0, \dots x_k) = Span(X) \cap S_n \setminus Cut(X) \qquad Aff(x_0, \dots x_k) = Span(X) \cap S_n$$

KBS/FBS: look at index of the Hessian of λ-variance

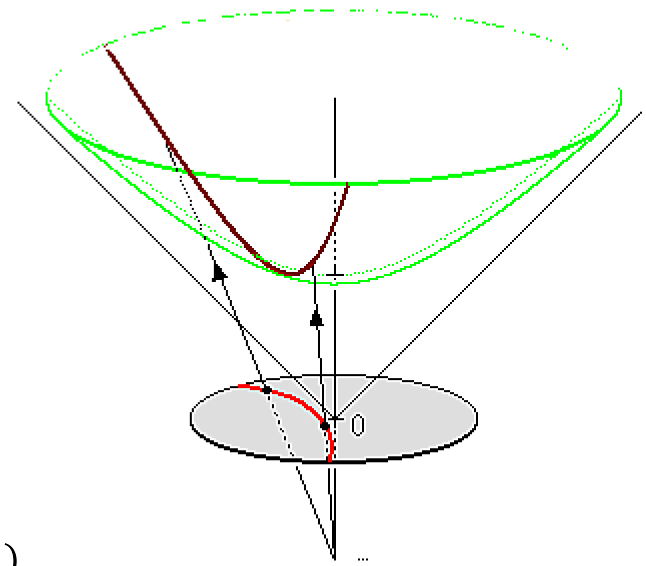
$$\mathrm{H}(\mathrm{x},\lambda) = \sum \lambda_i \theta_i \cot(\theta_i)\,(\mathrm{Id} - \mathrm{xx}^{\mathrm{t}}) + \sum (1 - \lambda_i \theta_i \cot(\theta_i))\, \overrightarrow{x x_i}\, \overrightarrow{x x_i}^{t}$$

- Complex algebric geometry problem [Buss & Fillmore, ACM TG 2001]
- 3 points of the n-sphere: EBS partitioned in cell complex by index of critical point
- **KBS/EBS less interesting than EBS/affine span**



Weighed Hessian index: **brown = -2 (min) = KBS** / green = -1 (saddle) / blue = 0 (max)

Example on the hyperbolic space



Manifold

- Unit pseudo-sphere $\mathcal{M} = \boldsymbol{H_n}$ embedded in Minkowski space $\mathbb{R}^{\mathbf{1},\boldsymbol{n}}$
- $\|x\|_*^2 = -x_0^2 + x_1^2 + \cdots x_n^2 = -1$

Exp and log map

$$\exp_x(v) = \cosh(\|v\|_*)\, x + \frac{\sinh(\|v\|_*)}{\|v\|_*} v$$

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

Distance $dist(x, y) = \|\log_x(y)\|_* = \theta$

Punctured hyperbolic space: no cut locus to exclude

Example on the hyperbolic space

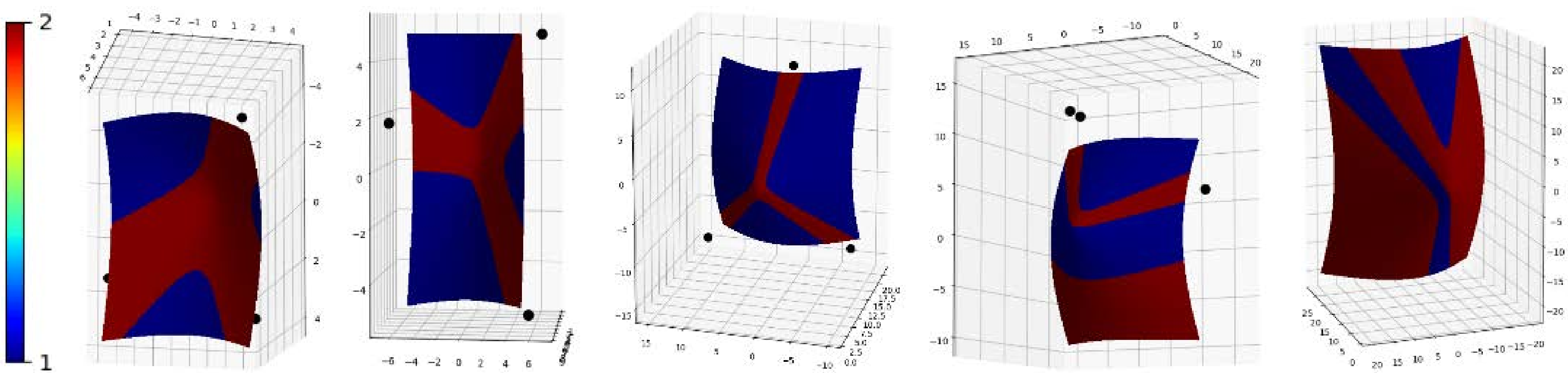
EBS = Affine span: great sub-hyperboloids spanned by reference points

$$\mathrm{EBS}(x_0, \ldots x_k) = Aff(x_0, \ldots x_k) = Span(X) \cap H_n$$

KBS: locus of maximal index of the Hessian of λ-variance

$$\mathrm{H}(\mathrm{x},\lambda) = \sum \lambda_i \theta_i \coth(J + J\mathrm{x}\mathrm{x}^{\mathrm{t}} J^t) + \sum (1 - \lambda_i \coth(\theta_i)) J\, \overrightarrow{xx_i}\, \overrightarrow{xx_i}^t J^t$$

- Complex algebric geometry problem
- 3 points on H^n: better than for spheres, but still disconnected components



Weighted Hessian Index: **brown = -2 (min) = KBS** / blue = 1 (saddle)

Geodesic subspaces are limit cases of affine span

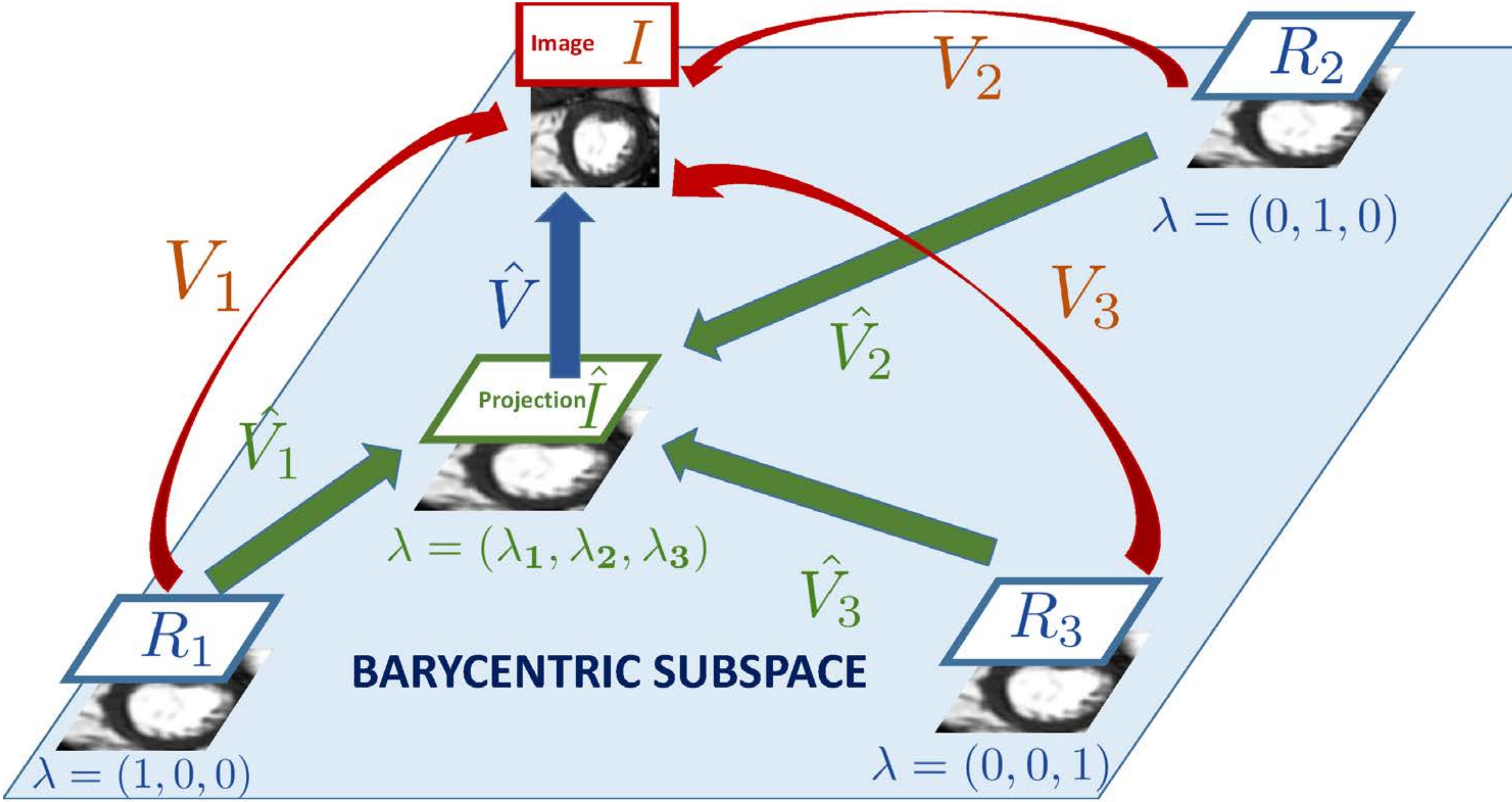
Theorem

- $GS(x, w_1, \dots w_k) = \{\exp_x(\sum_i \alpha_i w_i) \; for \; \alpha \in R^k \}$ is the limit of $Aff\left(x_0, \exp_{x_o}(\epsilon\, w_1), \dots \exp_{x_o}(\epsilon\, w_k)\right)$ when $\epsilon \to 0$.
- Reference points converge to a 1st order (k,n)-jet
  - PGA [Fletcher et al. 2004, Sommer et al. 2014]
  - GPGA [Huckemann et al. 2010]

Conjecture

- This can be generalized to higher order derivatives
  - Quadratic, cubic splines [Vialard, Singh, Niethammer]
  - Principle nested spheres [Jung, Dryden, Marron 2012]
  - Quotient of Lie group action [Huckemann, Hotz, Munk, 2010]

Application in Cardiac motion analysis

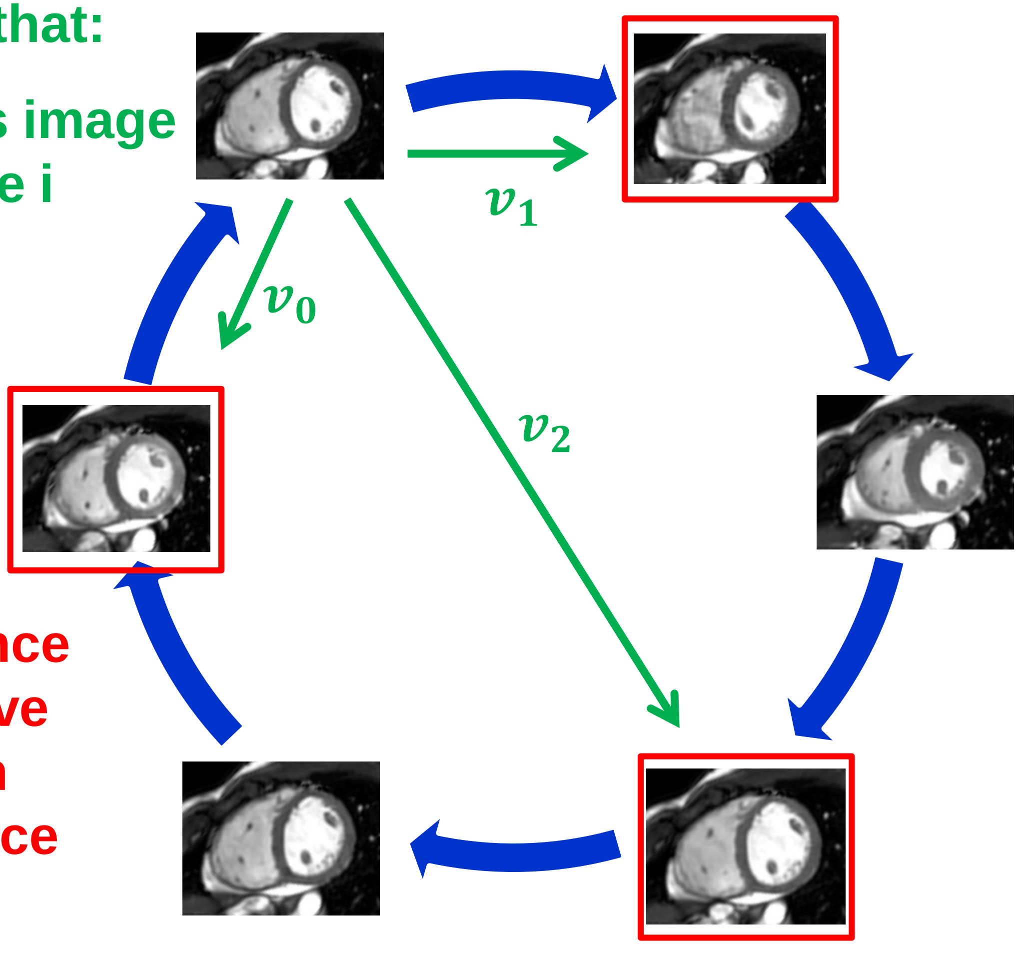


[Marc-Michel Rohé et al., MICCAI 2016, MedIA 45:1-12, 2018]

Application in Cardiac motion analysis

Find weights λ_i and SVFs v_i such that:

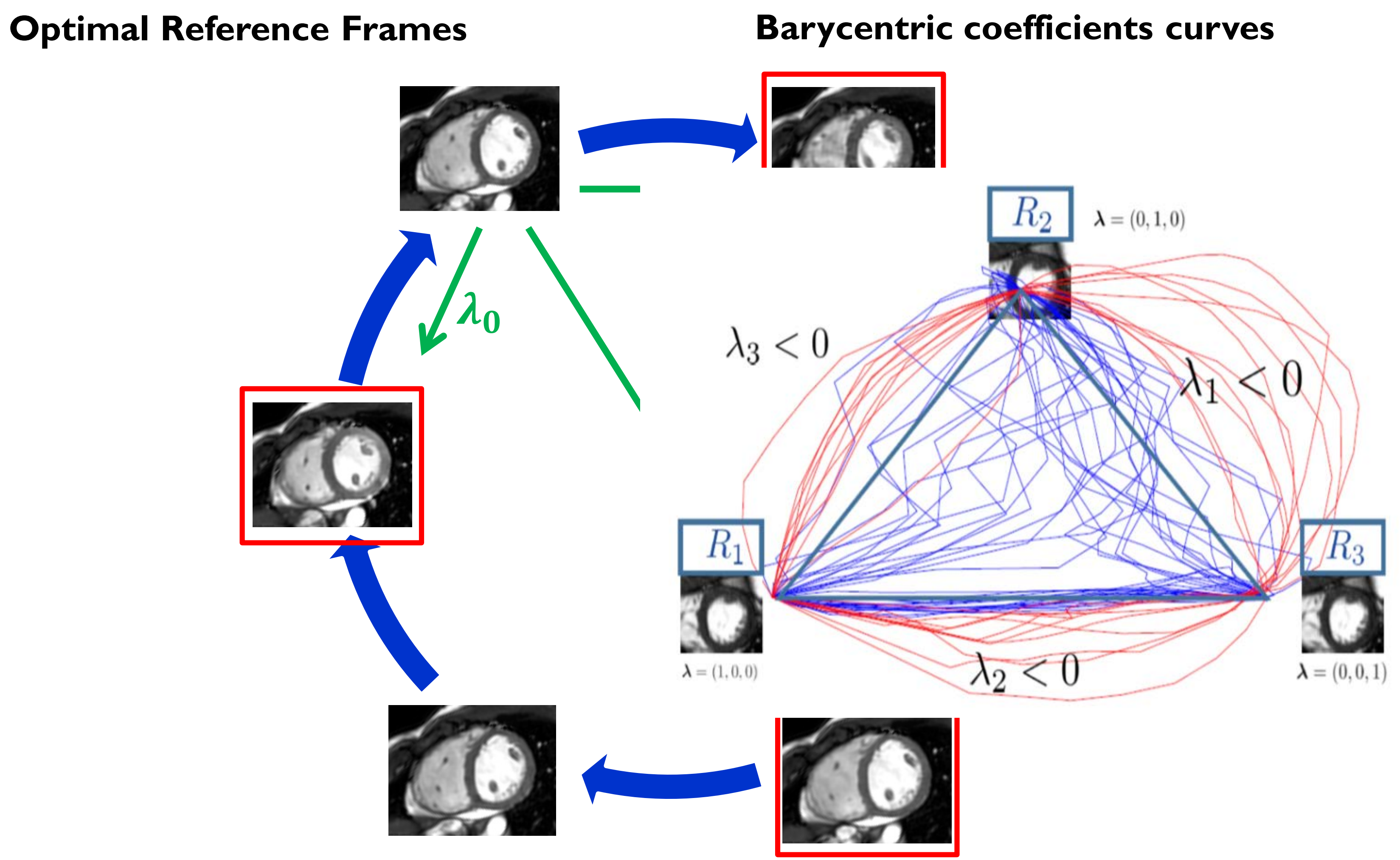
- v_i **registers image to reference i**
- $\sum_i \lambda_i v_i = 0$



Optimize reference images to achieve best registration over the sequence

[Marc-Michel Rohé et al., MICCAI 2016, MedIA 45:1-12, 2018]

Application in Cardiac motion analysis

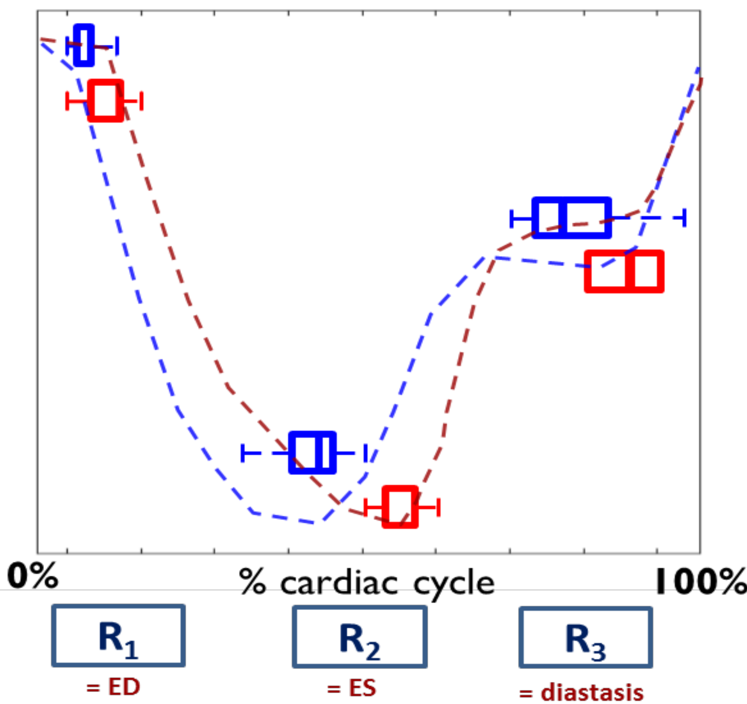


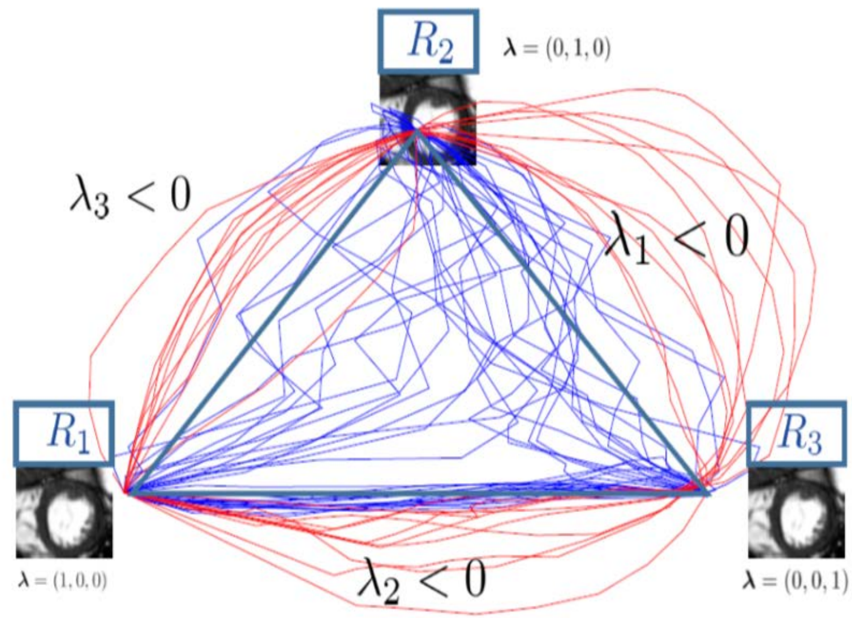
[Marc-Michel Rohé et al., MICCAI 2016, MedIA 45:1-12, 2018]

Cardiac Motion Signature

Low-dimensional representation of motion using:

Optimal Reference Frames | **Barycentric coefficients curves**





Dimension reduction from **+10M voxels** to **3 reference** frames + **60 coefficients**

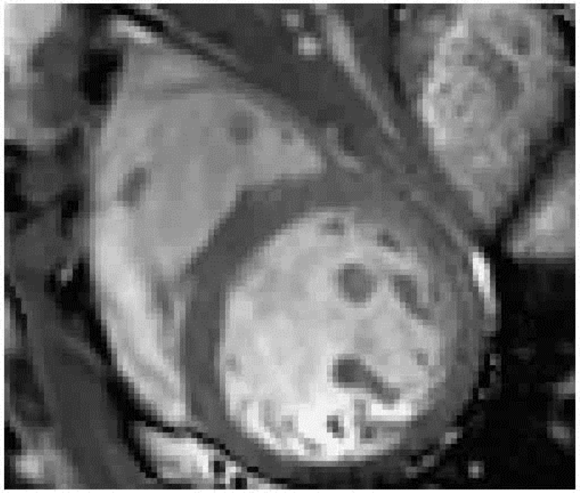
Tested on **10 controls** [1] and **16 Tetralogy of Fallot** patients [2]

[1] Tobon-Gomez, C., et al.: Benchmarking framework for myocardial tracking and deformation algorithms: an open access database. Medical Image Analysis (2013)

[2] Mcleod K., et al.: Spatio-Temporal Tensor Decomposition of a Polyaffine Motion Model for a Better Analysis of Pathological Left Ventricular Dynamics. IEEE TMI (2015)

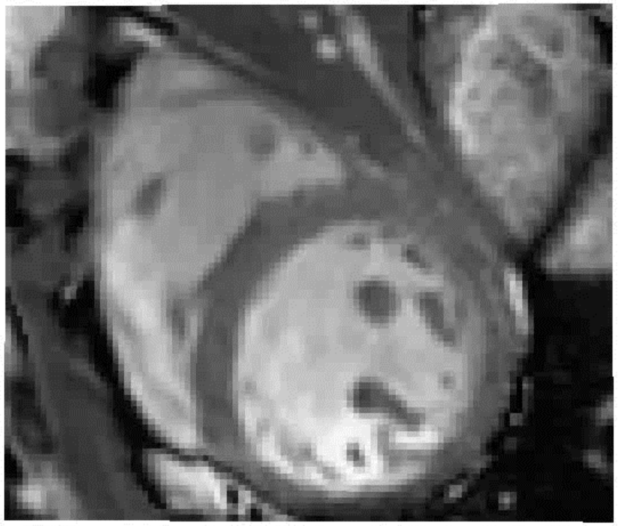
Cardiac motion synthesis

Original Sequence



30 images

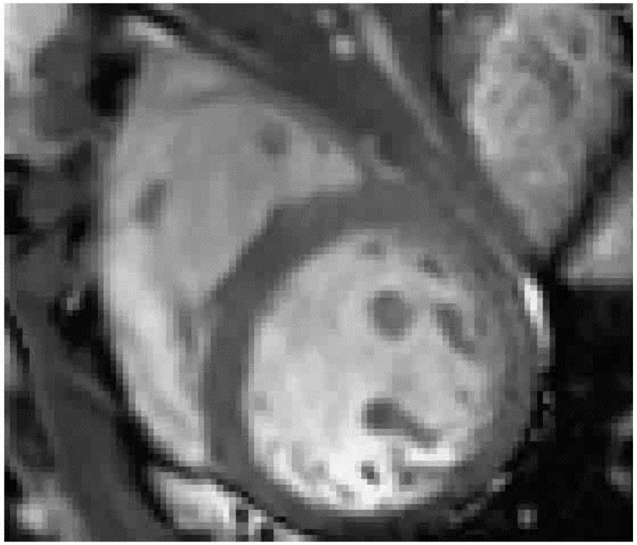
Barycentric Reconstruction (3 images)



3 images + 2 coeff.

Reconstr. error: 18.75
Compression ratio: 1/10

PCA Reconstruction (2 modes)



1 image + 2 SVF + 2 coeff.

Reconstr. error: 26.32 (+40%)
Compression ratio: 1/4

[Marc-Michel Rohé et al., MICCAI 2016, MedIA 45:1-12, 2018]

Geometric Statistics: Mathematical foundations and applications in computational anatomy

Intrinsic Statistics on Riemannian Manifolds

Manifold-Valued Image Processing

Metric and Affine Geometric Settings for Lie Groups

Parallel Transport to Analyze Longitudinal Deformations

Advances Statistics: CLT & PCA

- Estimation of the empirical Fréchet mean & CLT
- Principal component analysis in manifolds
- Natural subspaces in manifolds: barycentric subspaces
- Rephrasing PCA with flags of subspaces

Forward, Backward and Nested Analysis

Forward Barycentric Subspace (k-FBS) decomposition

- Iteratively add points x_j from j=0 to k
- $x_0 = Mean(y_j), \quad x_1 = argmin_x(\sigma^2_{out}(x_0, x)) \dots$ PGA-like
- Start with 2 points: $(x_0, x_1) = \mathrm{argmin}_{(x,y)}(\sigma^2_{out}(x, y))$ GPGA-like

Backward analysis: Pure Barycentric Subspace (k-PBS)

- Find $Aff(x_0, \dots x_k)$ minimizing the unexplained variance:

$$\sigma^2_{out}(x_0, \dots x_k) = \sum_j dist^2\left(y_j, Proj_{Aff(x_0 \dots x_k)}(y_j)\right)$$

- Iteratively remove one point from $(x_0, \dots x_j)$ from j=0 to k
- One optimization only for k+1 points and discrete backward reordering

From greedy to global optimization?

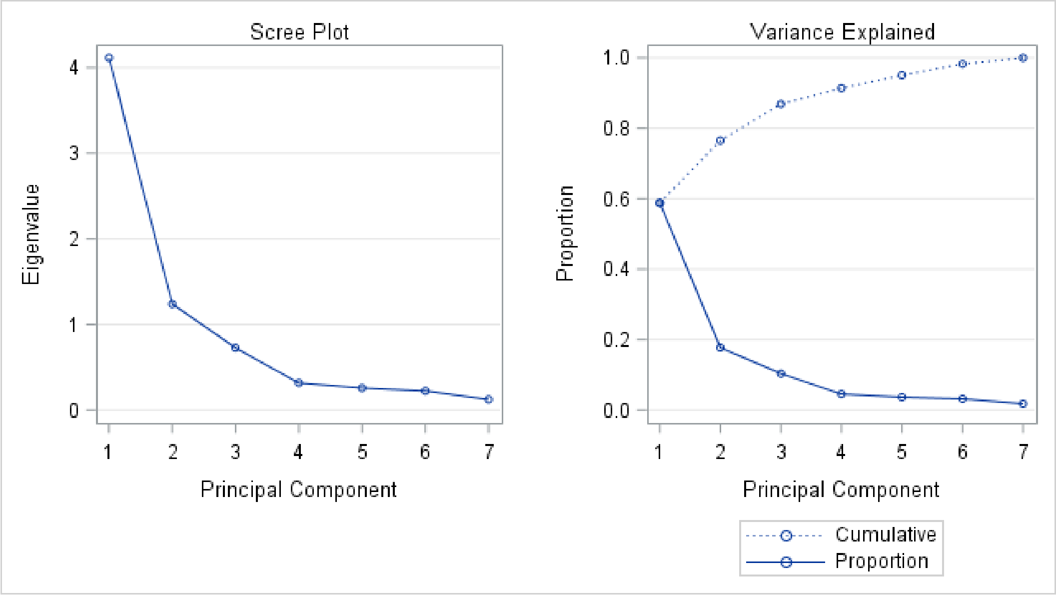
- Optimal unexplained variance → non nested subspaces
- Nested forward / backward procedures → not optimal
- Optimize first, decide dimension later → Nestedness required
  [Principal nested relations: Damon, Marron, JMIV 2014]

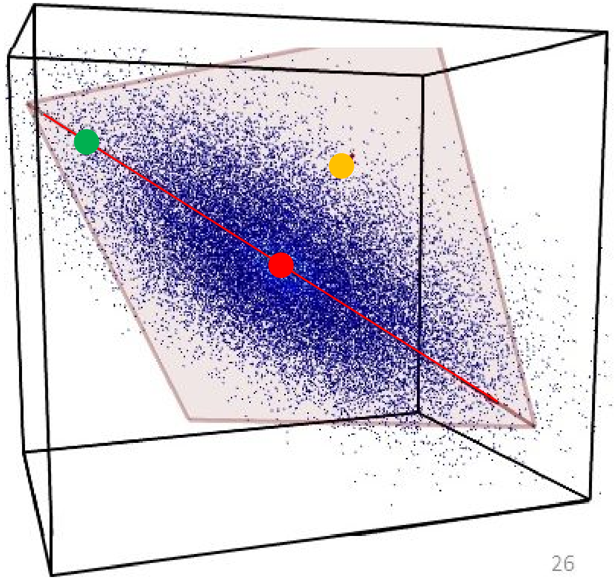
Barycentric Subspace Analysis (k-BSA)

The natural object for PCA: Flags of subspaces in manifolds

- $x_0 \prec x_1 \prec \cdots \prec x_k$ are k +1 n distinct ordered points of M.
- $FL(x_0 \prec x_1 \prec \cdots \prec x_k)$ is the sequence of properly nested subspaces $FL_{i(x_0 \prec x_1 \prec \cdots \prec x_k)} = Aff(x_0, \ldots x_i)$

$$Aff(x_0) = \{x_0\} \subset \ldots Aff(x_0, \ldots x_k) \ldots \subset Aff(x_0, \ldots x_n) = M$$
$$\sigma^2_{out}(x_0) \geq \ldots \geq \sigma^2_{out}(x_0, \ldots x_k) \geq \ldots \geq \sigma^2_{out}(x_0, \ldots x_n) = 0$$





Adapted from 3DM slides by Marc van Kreveld

Barycentric Subspace Analysis (k-BSA)

Accumulated unexplained variance (area under the curve)

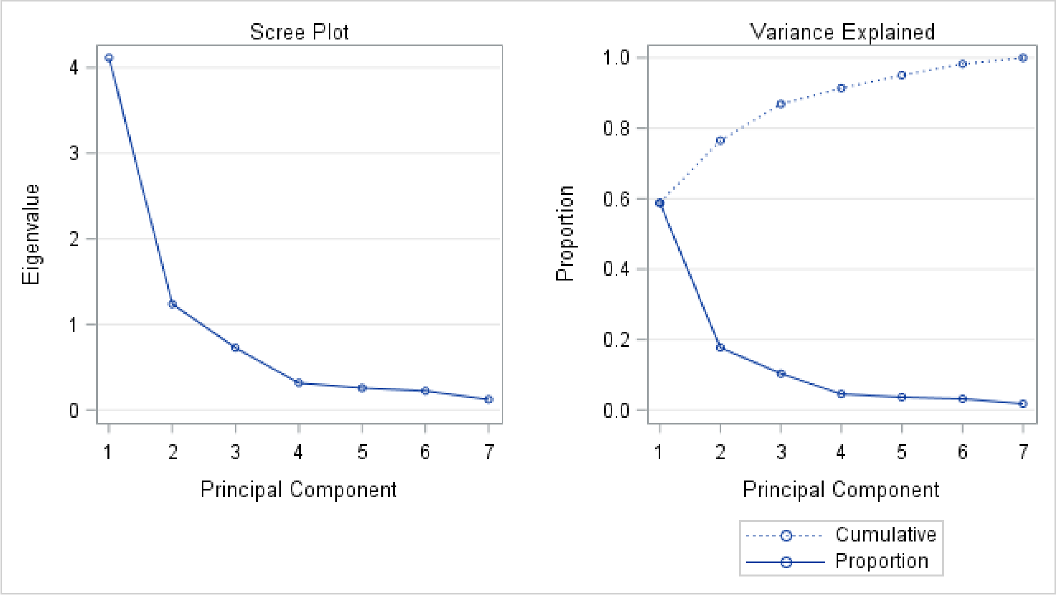
- k-BSA optimizes: $AUV(k) = \sum_{i=0}^{k} \sigma_{out}^2(x_0, \ldots x_i)$
- In a Euclidean space with Gaussian $N(x_0, \Sigma = diag(\sigma_1^2, \ldots \sigma_n^2))$

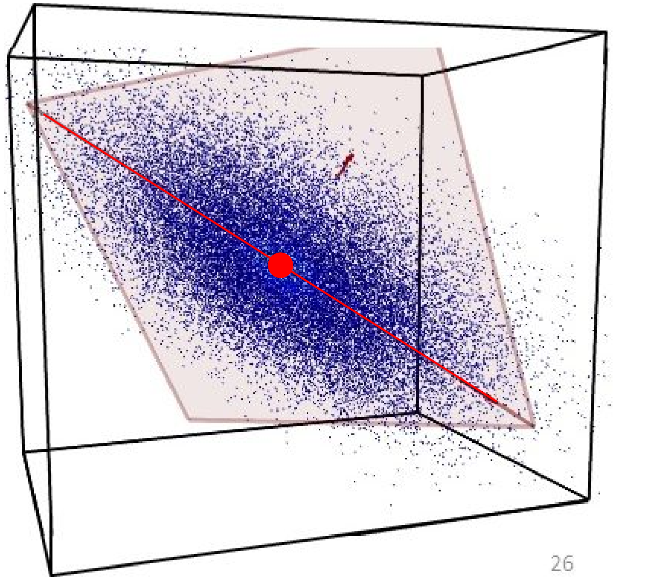
$\sigma_{out}^2(x_0, \ldots x_i) = \sigma_{i+1}^2 + \cdots \sigma_n^2 \rightarrow AUV(k) = \sum_{i=0}^{k} i\, \sigma_i^2 + (k+1) \sum_{i=k+1}^{n} \sigma_i^2$

→ minimal for ordered eigenmodes of Σ with $\sigma_1 \geq \sigma_2 \; \ldots \geq \sigma_n$

\

[Barycentric Subspace Analysis on Manifolds, Annals of Statistics 2018]



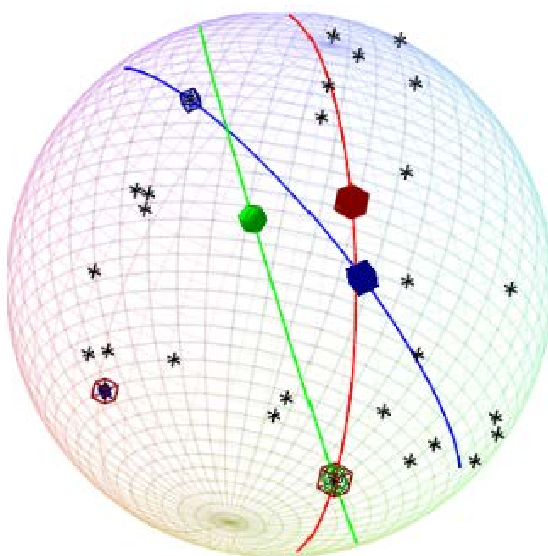


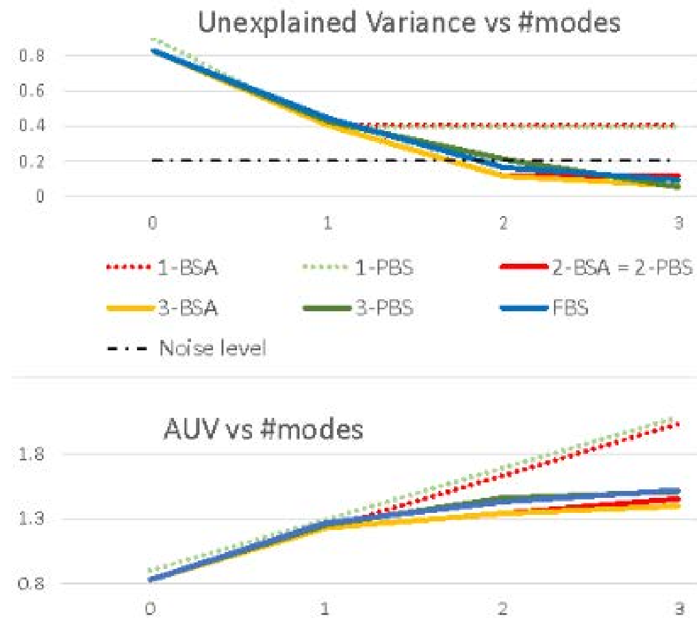
Adapted from 3DM slides by Marc van Kreveld

Sample-limited barycentric subspace inference

Restrict the inference to data points only

- Fréchet mean / template [Lepore et al 2008]
- First geodesic mode [Feragen et al. 2013, Zhai et al 2016]
- Higher orders: challenging with PGA… but not with BSA





- **FBS: Forward Barycentric Subspace**
- **k-PBS: Pure Barycentric Subspace with backward ordering**
- **k-BSA: Barycentric Subspace Analysis up to order k**

Robustness with L_p norms

Affine spans is stable to p-norms

- $\sigma^p(x,\lambda) = \frac{1}{p}\sum\lambda_i dist^p(x, x_i)/\sum\lambda_i$
- Critical points of $\sigma^p(x,\lambda)$ are also critical points of $\sigma^2(x,\lambda')$ with $\lambda'_i = \lambda_i\, dist^{p-2}(x, x_i)$ (non-linear reparameterization of affine span)

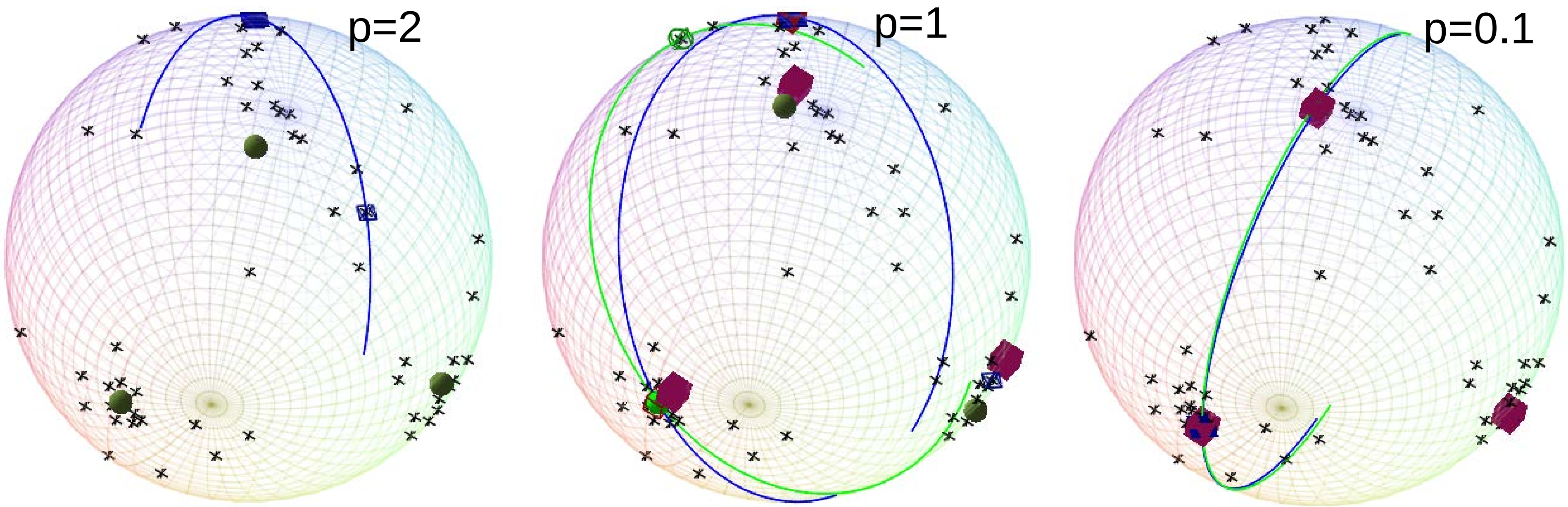
Unexplained p-variance of residuals

- $2 < p \rightarrow +\infty$: more weight on the tail,
  at the limit: penalizes the maximal distance to subspace
- $0 < p < 2$: less weight on the tail of the residual errors: statistically robust estimation
  - Non-convex for p<1 even in Euclidean space
  - But sample-limited algorithms do not need gradient information

Experiments on the sphere

3 clusters on a 5D sphere

- 10, 9 and 8 points (stddev 6 deg) around three orthogonal axes plus 30 points uniformly samples on 5D sphere

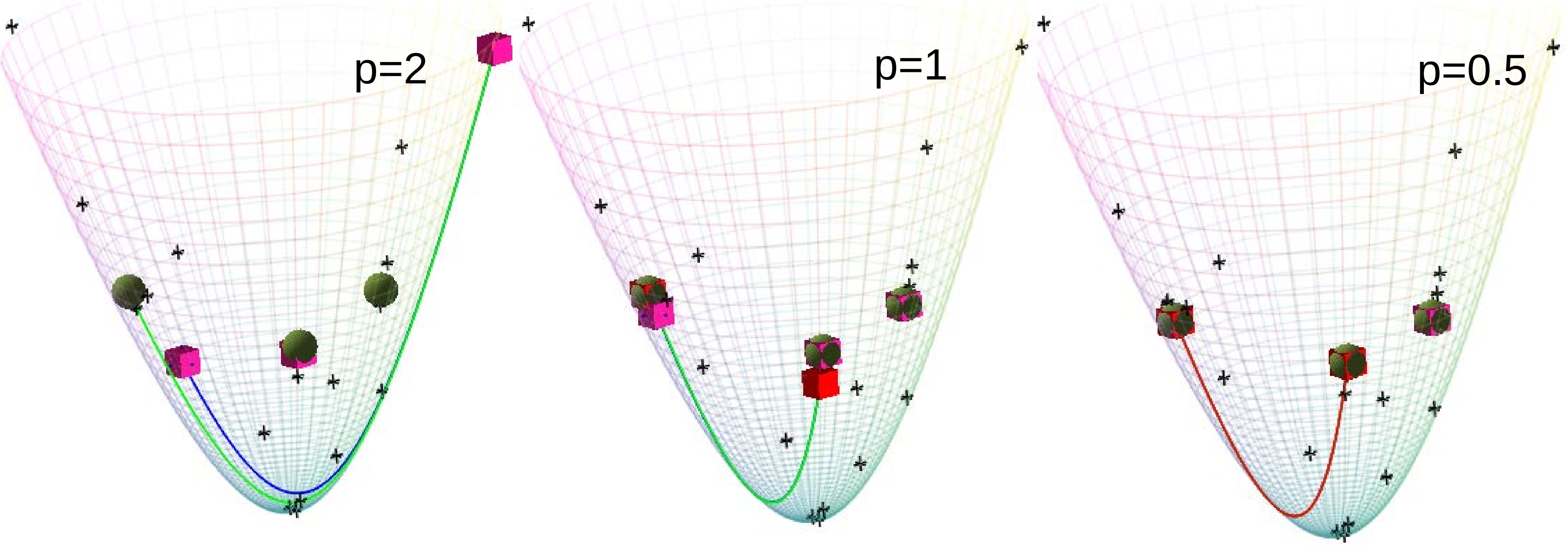


- **FBS: Forward Barycentric Subspace: mean and median not in clusters**
- **1-PBS / 2-PBS: Pure Barycentric Subspace with backward ordering: ok for k=2 only**
- **1-BSA / 2-BSA: Barycentric Subspace Analysis up to order k: less sensitive to p & k**

Experiments on the hyperbolic space

3 clusters on a 5D hyperboloid (50% outliers)

- 15 random points (stddev 0.015) around an equilateral triangle of length 1.57 plus 15 points of stddev 1.0 (truncated at max 1.5)



- **FBS: Forward Barycentric Subspace: ok for $p \leq 0.5$**
- **1-PBS / 2-PBS: Pure Barycentric Subspace with backward ordering: ok for k=2 only**
- **1-BSA / 2-BSA: Barycentric Subspace Analysis up to order k: ok for $p \leq 1$**

Take home messages

Natural subspaces in manifolds

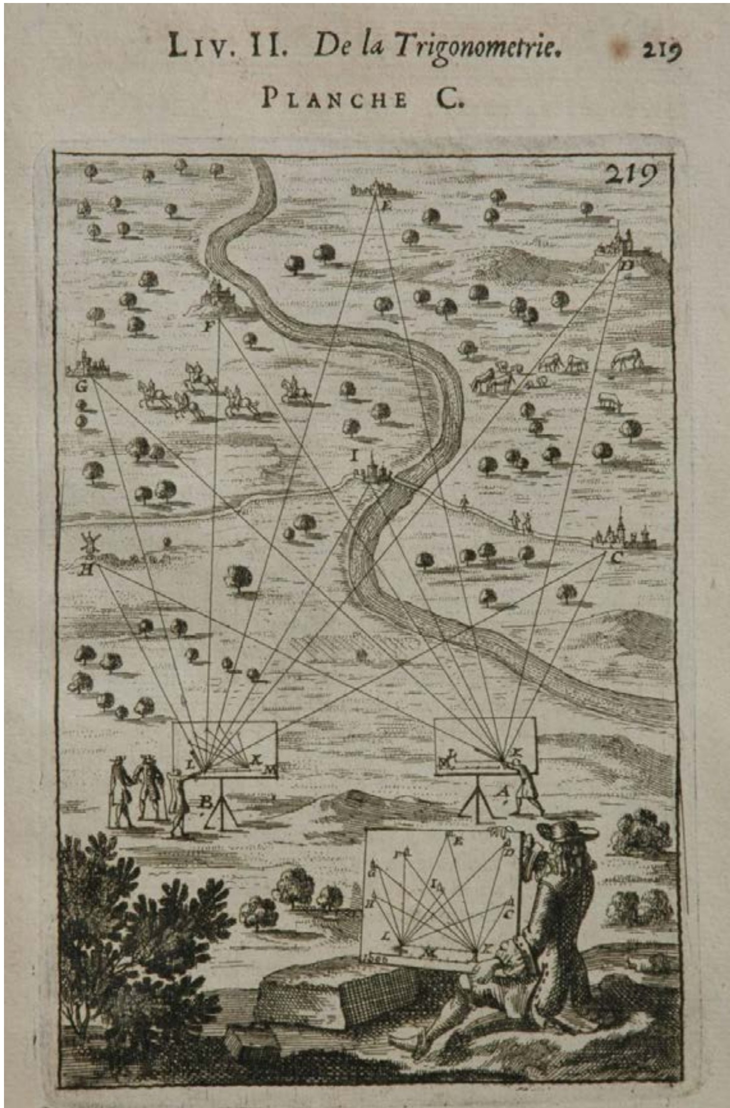
- PGA & Godesic subspaces:
  look at data points from the (unique) mean
- Barycentric subspaces:
  « triangulate » several reference points
  - Justification of multi-atlases?

Critical points (affine span) rather than minima (FBS/KBS)

- Barycentric coordinates need not be positive (convexity is a problem)
- Affine notion (more general than metric)
  - Generalization to Lie groups (SVFs)?

Natural flag structure for PCA

- Hierarchically embedded approximation subspaces to summarize / describe data



A. Manesson-Mallet. La géométrie Pratique, 1702

Open research avenues

Other iterative least squares methods?

- ICA, PLS
- Manifold learning → Submanifold learning

Modulate BSA to account for within subspace distribution

- Gaussian: central points
- Clusters: mixtures of modes
- Extremal references: archetypal analysis

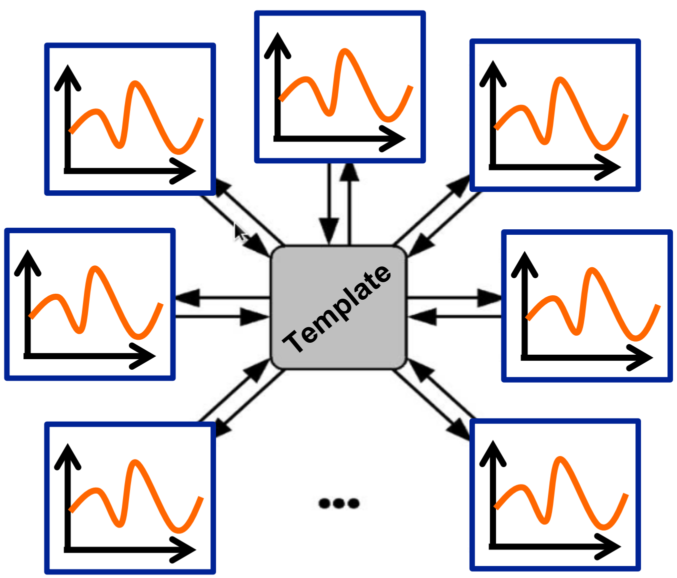
And applications

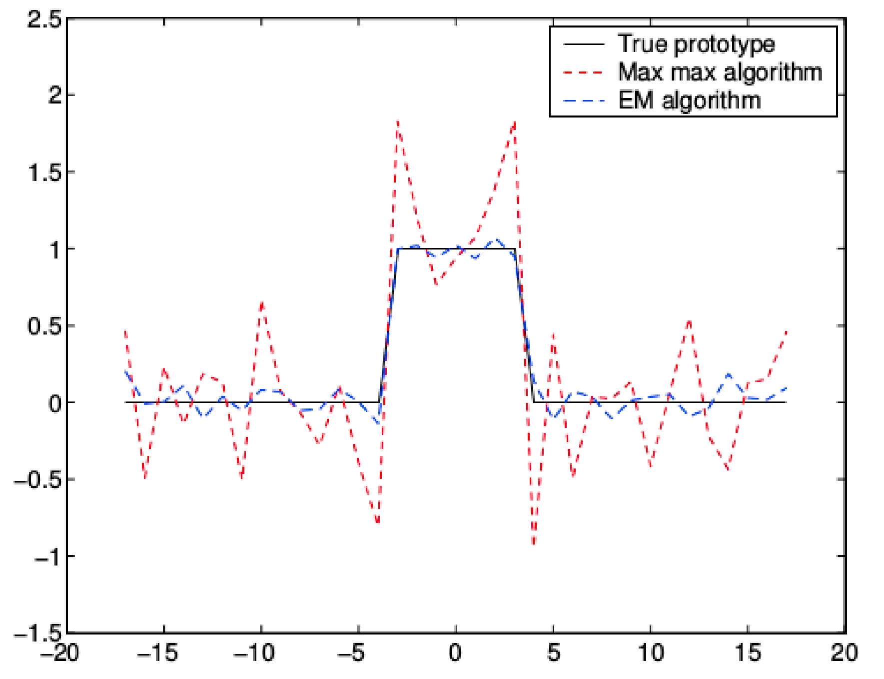
- Multi-atlases (brains, heart motion image sequences)
- SPD matrices (BCI)

Quotient spaces

Functions/Images modulo time/space parameterization

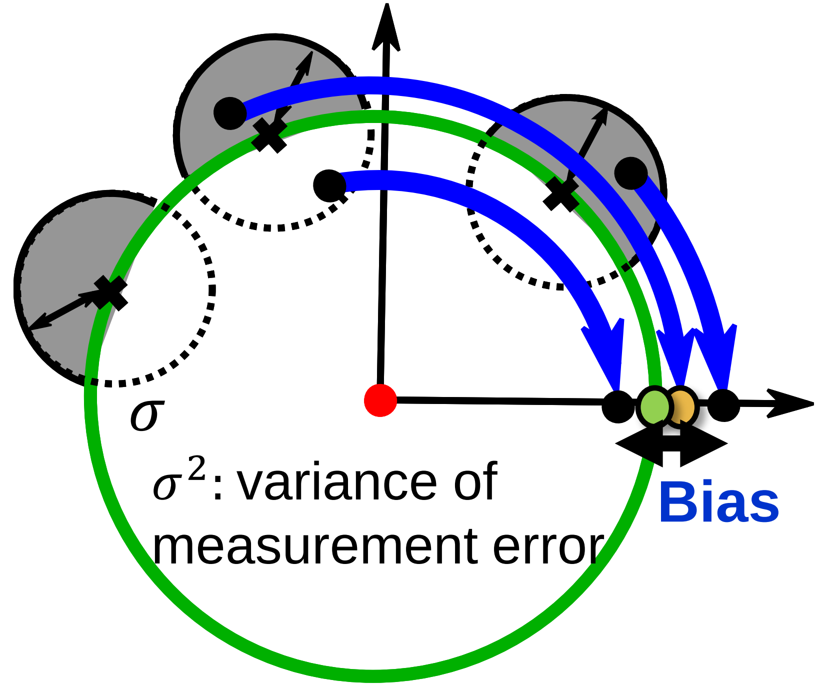
- Amplitude and phase discrimination problem





[Allassoniere, Amit, Trouvé, 2005],
Example by Loic Devillier, IPMI 2017

Noise in top space = Bias in quotient spaces



The curvature of the **template shape's orbit and presence of noise** creates a repulsive bias

Theorem [Miolane et al. (2016)]: Bias of estimator $\hat{T}$ of the template T

$$\text{Bias}(\hat{T}, T) = \frac{\sigma^2}{2} \boldsymbol{H}(T) + \mathcal{O}(\sigma^4)$$

where $\mathbf{H}(T)$: **mean curvature vector of template's orbit**

Extension to Hilbert of ∞-dim: bias for $\sigma > 0$, asymptotic for $\sigma \to \infty$,
[Devilliers, Allassonnière, Trouvé and XP. SIIMS 2017, Entropy, 2017]

→ Estimated atlas is topologically more complex than should be

References on Barycentric Subpsace Analysis

- **Barycentric Subspace Analysis on Manifolds**
  X. P. Annals of Statistics. 46(6A):2711-2746, 2018. [arXiv:1607.02833]
  - **Barycentric Subspaces and Affine Spans in Manifolds** Geometric Science of Information GSI'2015, Oct 2015, Palaiseau, France. LNCS 9389, pp.12-21, 2015.
    Warning: change of denomination since this paper: EBS →affine span
  - **Barycentric Subspaces Analysis on Spheres** Mathematical Foundations of Computational Anatomy (MFCA'15), Oct 2015, Munich, Germany. pp.71-82, 2015. https://hal.inria.fr/hal-01203815
- **Sample-limited L p Barycentric Subspace Analysis on Constant Curvature Spaces.** X.P. Geometric Sciences of Information (GSI 2017), Nov 2017, Paris, France. LNCS 10589, pp.20-28, 2017.
- **Low-Dimensional Representation of Cardiac Motion Using Barycentric Subspaces: a New Group-Wise Paradigm for Estimation, Analysis, and Reconstruction.** M.M Rohé, M. Sermesant and X.P. Medical Image Analysis vol 45, Elsevier, April 2018, 45, pp.1-12.
  - **Barycentric subspace analysis: a new symmetric group-wise paradigm for cardiac motion tracking.** M.M Rohé, M. Sermesant and X.P. Proc of MICCAI 2016, Athens, LNCS 9902, p.300-307, Oct 2016.