Riemannian Geometry Methods and Tools for EEG preprocessing, analysis and classification

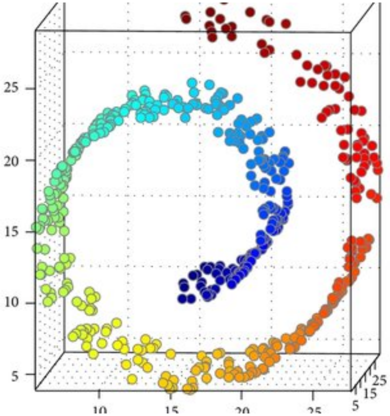
Why Riemannian Geometry works so well?

Florian YGER
LAMSADE - MILES

Université Paris-Dauphine, PSL Research University, LAMSADE, CNRS

Foreword

What this presentation is not about

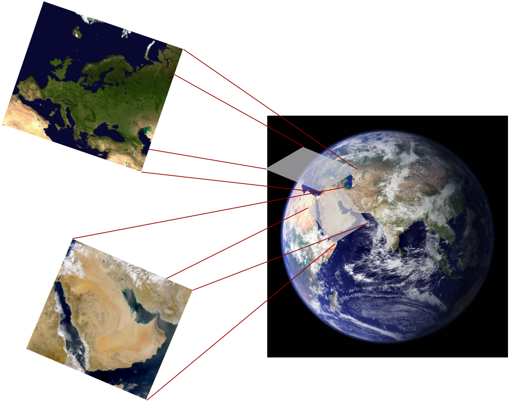


Manifold learning (non-linear dimensionality reduction)

- looking for a subspace where the data are distributed
- in the sequel, the manifold is known (and handled as a constraint).

Some technicalities about Riemannian geometry

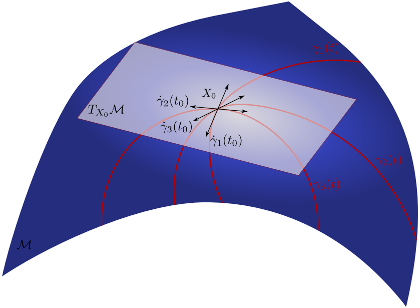
A well-known manifold



Intuitions :

- "curved" subspace of $\mathbb{R}^N$
- locally approximated by hyperplanes (i.e. maps)
- from one point to another through geodesics

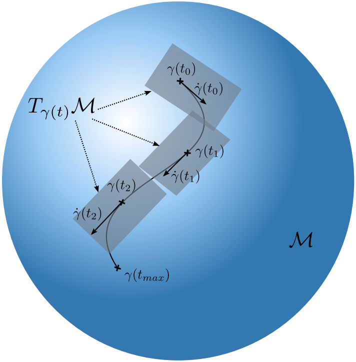
Local approximation



Tangent space $T_{X_0}\mathcal{M}$
For $\mathcal{M}$, at X_0 :

- the set of gradients at X_0 of every curve $\gamma_i(t)$ passing through this point (tangent plane)
- equipped with a scalar product (riemannian metric).

Riemannian distance



- the lenght of a curve is deduced by integrating the norm of its gradient in the tangent spaces

$$L_g(\gamma) = \int_a^b ||\gamma'(t)||_g dt$$

- depends on the chosen geometry (i.e. the riemannian metric)

Exp/Log map

Exponential map
The exponential map sends points from a given tangent plan to the manifold. $\gamma_v(t)$ is the geodesic from p with initial acceleration v

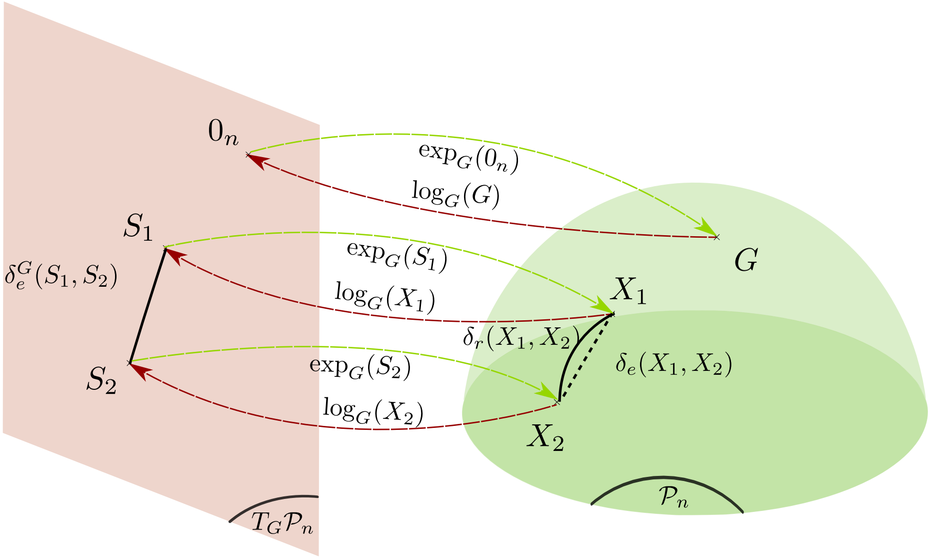
$$T_p\mathcal{M} \to \mathcal{M}$$
$$\exp_p(v) = \gamma_v(1)$$

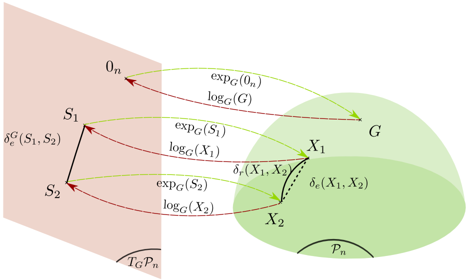
Logarithmic map
The logarithmic map flattens the manifold around one point (i.e. the tangent plan tangent at this point).

Remarks

- back and forth between a manifold $\mathcal{M}$ and (euclidean) $T_p\mathcal{M}$
- usually computationally expensive operations (retractions).

Goings and comings between a manifold and a tangent plane





A word of caution

Applying linearisation implies a deformation (which depends on the curvature of the space and of the distance to from the reference point)

A small recap

What do we have now ?

- an elegant way to model some smooth constraints on the data.
- with the proper metric, geodesics allow to search the space of feasible solutions.
- linearisation (with Exp/Log or an approx.) offers practical tools

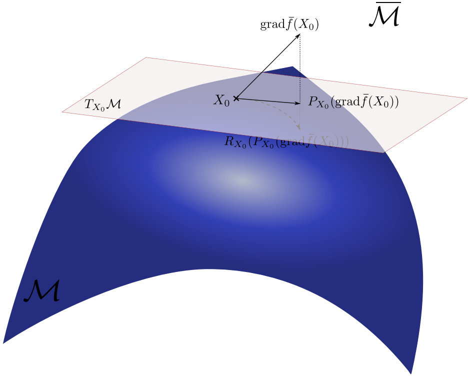
What do we need ?

- an optimization scheme.

Riemannian optimization in one slide

Key step

1. (euclidean) gradient
2. projection on the tangent plan (riemannian gradient)
3. exponential map (or retraction)



Philosophy of this approach

Move on a geodesic in the direction of the gradient (i.e. geodesic minimising the cost).

A notion of convexity

Geodesic convexity

- g-convexity is defined as (with $\gamma(0) = x$ et $\gamma(1) = y$) :

$$f(\gamma(t)) \leq (1 - t)f(x) + tf(y).$$

- a g-convex function has a unique minimum (reachable through a Riemannian descent method)

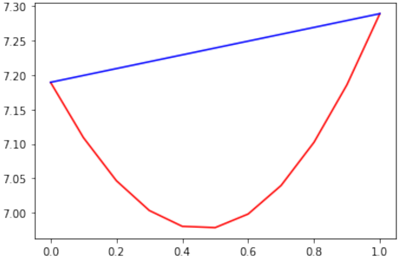
A notion of convexity

Geodesic convexity

- g-convexity is defined as (with $\gamma(0) = x$ et $\gamma(1) = y$) :

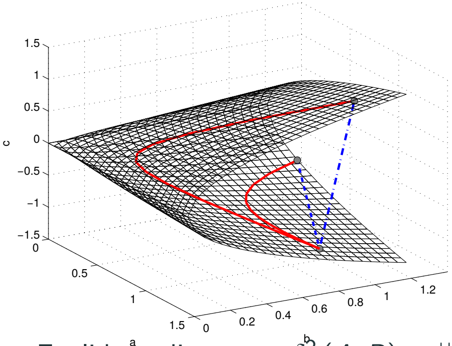
$$f(\gamma(t)) \leq (1-t)f(x) + tf(y).$$

- a g-convex function has a unique minimum (reachable through a Riemannian descent method)



A manifold of interest for EEG data

(Strictly) definite-positive matrices

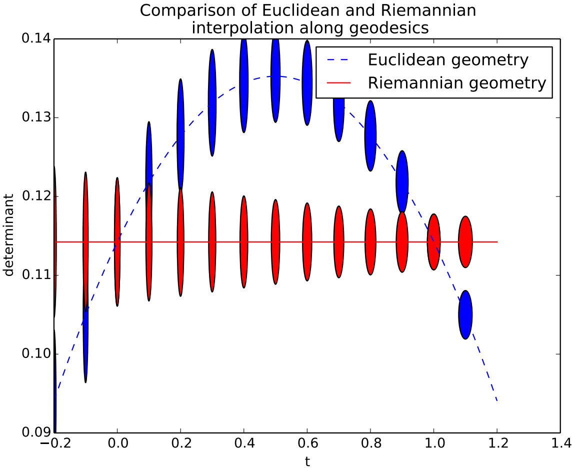


$$C = \begin{vmatrix} a & b \\ b & c \end{vmatrix}$$

$$ac - b^2 > 0$$

- Euclidean distance : $\delta_E^2(A, B) = ||A - B||_{\mathcal{F}}^2$
  interpolation is possible but to the cost of the *swelling effect*.
- Riemannian distance (AIRM) :
  $\delta_R^2(A, B) = ||\log(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})||_{\mathcal{F}}^2$.
  interpolation and extrapolation without any *swelling effect*.
- LogEuclidean distance : $\delta_L^2(A, B)_R = ||\log_R(A) - \log_R(B)||_{\mathcal{F}}^2$

Swelling effect



It can occur that $det(\frac{A+B}{2}) \geq \max(det(A), det(B))$, which is an artifact of the Euclidean framework.

Where do we find those matrices ?

Covariance-based features

$X \in \mathbb{R}^{n \times s}$ a an epoch of signal and $T \in \mathbb{R}^{n \times s}$ a template

- spatial covariance matrix: $C_s = \frac{1}{s} X X^\top$ - with the variance/power of electrodes on the diagonal,
- template-signal covariance: $C_T = \begin{pmatrix} TT^\top & TX^\top \\ XT^\top & XX^\top \end{pmatrix}$
- filtered signal covariance: $C_f = \begin{pmatrix} X_{f_1} X_{f_1}^\top & \cdots & X_{f_1} X_{f_F}^\top \\ \vdots & \ddots & \vdots \\ X_{f_F} X_{f_1}^\top & \cdots & X_{f_F} X_{f_F}^\top \end{pmatrix}$

  with the X_f filtered versions of the original signal.

Simple yet efficient approach I

MDRM

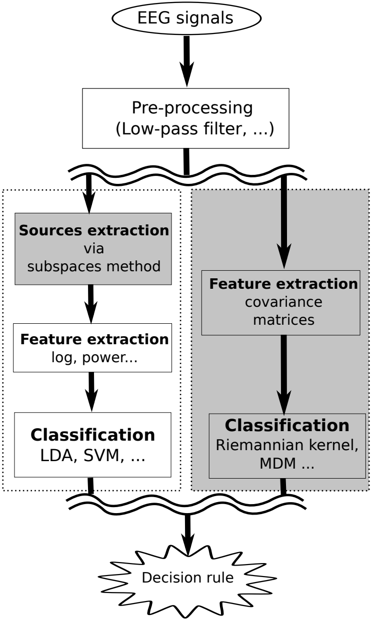
- algorithm
  For each class j $\quad \min_{X_j \in \mathcal{P}_n} \sum_{i \in \text{Class } j} \delta_R^2(C_i, X_j)$
- j independent problems of Fréchet/Karcher averaging
- prediction for C: $argmin_j \delta_R^2(C, X_j)$

Simple yet efficient approach II

TSLDA

- algorithm
  compute M the Fréchet mean of the whole dataset of covariances
  map each covariance C_i to S_i in $T_G\mathcal{P}_n$ with $\log_M(C_i) = \log(M^{-\frac{1}{2}} C_i M^{-\frac{1}{2}})$
  apply a simple LDA on the extracted feature
- interpretation in term of withening for $M^{-\frac{1}{2}} C_i M^{-\frac{1}{2}}$
- linearization of the manifold around M

A simpler framework



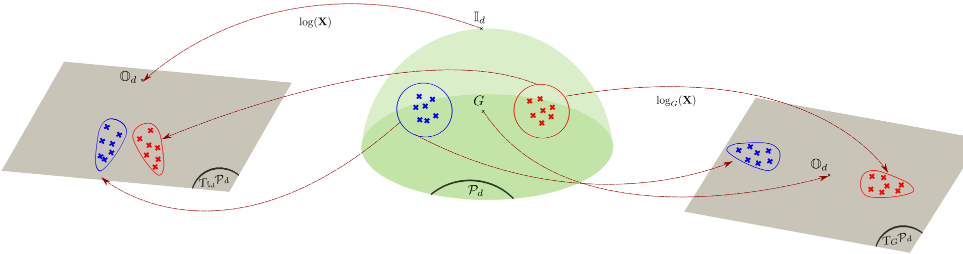
Questions raised by those approaches

Limits of TSLDA

- linearizing a manifold implies a deformation (when the mapped data are far from the origin of the mapping)
- the choice of reference point (Riemannian mean) is thought to minimize the distortion over the dataset
- using the AIRM geometry, the reference point of a tangent space also induces a metric

$$\forall A, B \in T_M\mathcal{P}_n \quad \langle A, B\rangle_M = \operatorname{tr}\left(M^{-1}AM^{-1}B\right).$$

Implied deformations



An interesting property

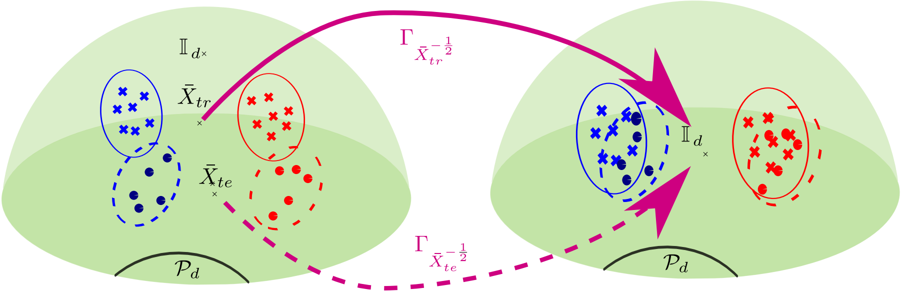
Affine invariance
Let us define the following *congruent transform*, for any $C, G \in \mathcal{P}_n$

$$\Gamma_{G^{-\frac{1}{2}}}(C) = G^{-\frac{1}{2}} C G^{-\frac{1}{2}}$$

It could be a permutation matrix (shuffling the electrodes) or a mixing matrix (mixing the sources).

- invariance of δ_R to this congruent transform/whitening $\delta_R(A, B) = \delta_R(\Gamma_W(A), \Gamma_W(B))$
- impact on the logEuclidean distance (depending on the reference point chosen)

Serendipitious observation about non-stationarity on manifold



- affine-invariance of δ_R to $\Gamma_W(.)$
  i.e.$\delta_R(A, B) = \delta_R(\Gamma_W(A), \Gamma_W(B))$
- between two sessions, sometimes a different whitening Γ seems to occur.

Conclusion

Finally

Why does it work so well ?

- simple framework with few parameters (avoiding overfitting)
- unaffected by the swelling effect
- natively incorporate invariances (e.g. affine-invariance)
- g-convexity in some well-posed problems (such as the Karcher averaging)

A word a caution

- curvature and kernels do not mix well (and kernels are not so easy to build in this context)
- rank-deficiency is a real issue
- performance may degrade in high dimension.

Perspectives

On the edge

- Apply RG to other DP matrices (e.g. from connectivity)
- use the RG to other paradigm and tasks
- transfer learning using RG
- leverage missing data and dimensionality reduction using RG
- investigate the use of other manifolds

Perspectives

On the edge

- Apply RG to other DP matrices (e.g. from connectivity)
  RIGOLETTO : joint work with MC. Corsi & S. Chevallier - *Riemannian Geometry on Connectivity for Clinical BCI*, ICASSP 2020 & winning method of BCI challenge WCCI
- use the RG to other paradigm and tasks
- transfer learning using RG
- leverage missing data and dimensionality reduction using RG
- investigate the use of other manifolds

Perspectives

On the edge

- Apply RG to other DP matrices (e.g. from connectivity)
- use the RG to other paradigm and tasks (e.g. workload)
- transfer learning using RG
- leverage missing data and dimensionality reduction using RG
- investigate the use of other manifolds

Perspectives

On the edge

- Apply RG to other DP matrices (e.g. from connectivity)
- use the RG to other paradigm and tasks
- transfer learning using RG
  see : *Dimensionality transcending: a method for merging BCI datasets with different dimensionalities* by P. Rodrigues, M. Congedo & C Jutten, TBME 2020
- leverage missing data and dimensionality reduction using RG
- investigate the use of other manifolds

Perspectives

On the edge

- Apply RG to other DP matrices (e.g. from connectivity)
- use the RG to other paradigm and tasks
- transfer learning using RG
- leverage missing data and dimensionality reduction using RG
  joint work with I. Horev & M. Sugiyama - *Geometry-aware principal component analysis for symmetric positive definite matrices*, ACML 2015
  joint work with S. Chevallier, Q. Barthelemy & S. Sra - *Geodesically-convex optimization for averaging partially observed covariance matrices*, ACML 2020
- investigate the use of other manifolds

Perspectives

On the edge

- Apply RG to other DP matrices (e.g. from connectivity)
- use the RG to other paradigm and tasks
- transfer learning using RG
- leverage missing data and dimensionality reduction using RG
- investigate the use of other manifolds (e.g. Grassmann manifold for subspaces or fixed-rank matrix manifold)

Wrap up

Take home message

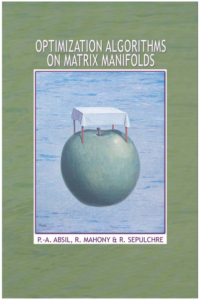
- Riemannian geometry as a way to incorporate prior knowledge on data/structural constraints
- Rich theory but easily applicable (given the available libraries)

What's next ?

- from metric learning on non-Euclidean data to deep models
  1. Geometric Deep Learning: going beyond Euclidean Data, Bronstein et al., IEEE Sig. Proc. Magazine, 2017
  2. A Riemannian Network for SPD Matrix Learning, AAAI 2017
- potential methodological pitfall (scarse data for deep models)

Some core references on manifolds

1. Edelman, A., Arias, T.A. and Smith, S.T., 1998. **The geometry of algorithms with orthogonality constraints**. SIAM journal on Matrix Analysis and Applications.
2. Absil, P.A., Mahony, R. and Sepulchre, R., 2009. **Optimization algorithms on matrix manifolds**. Princeton Univ. Press.
3. Boumal, N., 2020. **An introduction to optimization on smooth manifolds**.



Some core references on manifolds

1. Edelman, A., Arias, T.A. and Smith, S.T., 1998. **The geometry of algorithms with orthogonality constraints**. SIAM journal on Matrix Analysis and Applications.
2. Absil, P.A., Mahony, R. and Sepulchre, R., 2009. **Optimization algorithms on matrix manifolds**. Princeton Univ. Press.
3. Boumal, N., 2020. **An introduction to optimization on smooth manifolds**.