ON THE USE OF GEODESIC TRIANGLES BETWEEN GAUSSIAN DISTRIBUTIONS FOR CLASSIFICATION PROBLEMS

IEEE ICASSP 2022 - Singapore

A. Collas[1], F. Bouchard[2], G. Ginolhac[3], A. Breloy[4], C. Ren[1], J.-P. Ovarlez[1,5]

[1]SONDRA, CentraleSupélec, Univ. Paris-Saclay
[2]CNRS, L2S, CentraleSupélec, Univ. Paris-Saclay
[3]LISTIC, Univ. Savoie Mont Blanc
[4]LEME, Univ. Paris Nanterre
[5]DEMR, ONERA, Univ. Paris-Saclay

E-mail: antoine.collas@centralesupelec.fr

Time series in remote sensing and classification

Time series in remote sensing

In recent years, many image time series have been taken from the **earth** with different technologies: **SAR**, **multi/hyper spectral imaging**, ...

Objective

Segment semantically these data using **spatial** information, **temporal** information and **sensor diversity** (spectral bands, polarization...).

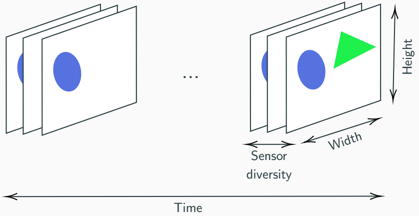


Figure 1: Multivariate image time series.

Applications

Disaster assessment, activity monitoring, land cover mapping, crop type mapping, ...

Example of multi-spectral time series

Breizhcrops dataset[1] [1]:

- more than 600 000 crop time series across the whole Brittany,
- 13 spectral bands, 9 classes.

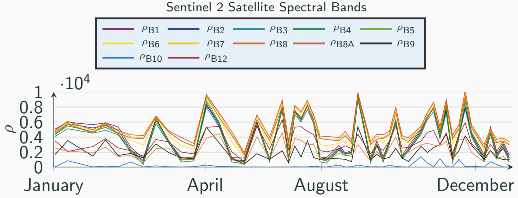


Figure 2: Reflectances ρ of a time series of **meadows**.

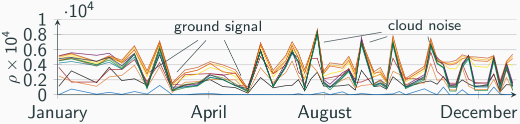


Figure 3: Reflectances ρ of a time series of **corn**.

[1] https://breizhcrops.org/

Clustering/classification pipeline and Riemannian geometry

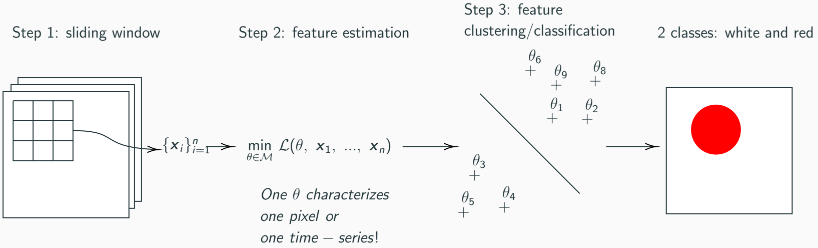


Figure 4: Clustering/classification pipeline.

Examples of θ:

$\theta = \Sigma$ a covariance matrix, $\theta = (\boldsymbol{\mu}, \Sigma)$ a vector and a covariance matrix, ...

Riemannian geometry and optimization

What is a Riemannian manifold ?

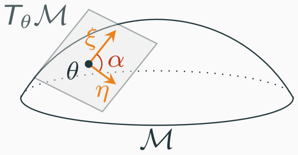


Figure 5: A Riemannian manifold.

Curvature induced by:

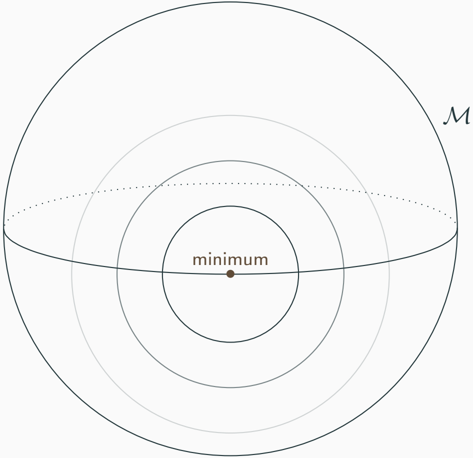
- constraints, *e.g.* the sphere: $\|\boldsymbol{x}\| = 1$,
- the Riemannian metric, *e.g.* on $\mathcal{S}_p^{++}$: $\langle \boldsymbol{\xi}_\Sigma, \boldsymbol{\eta}_\Sigma \rangle_\Sigma^{\text{FIM}} = \text{Tr}\left(\Sigma^{-1}\boldsymbol{\xi}_\Sigma\Sigma^{-1}\boldsymbol{\eta}_\Sigma\right)$.

Examples of Riemannian manifolds $\mathcal{M}$:

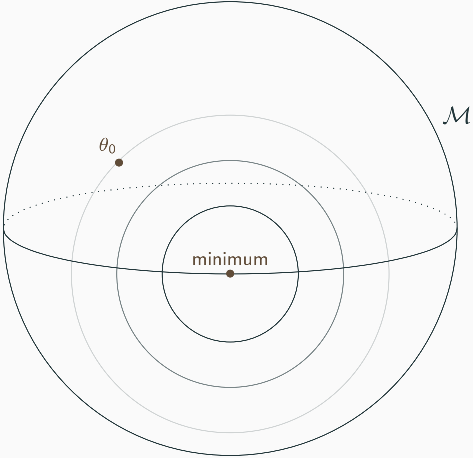
- linear space (no constraints): $\mathbb{R}^{p\times p}$
- orthogonality constraints: $\text{St}_{p,k} = \{\boldsymbol{U} \in \mathbb{R}^{p\times k} : \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_k\}$
- positivity constraints: $\mathcal{S}_p^{++} = \{\Sigma \in \mathcal{S}_p : \forall \boldsymbol{x} \neq \boldsymbol{0} \in \mathbb{R}^p,\ \boldsymbol{x}^T\Sigma\boldsymbol{x} > 0\}$
- norm constraints: $S^{p^2-1} = \{\boldsymbol{X} \in \mathbb{R}^{p\times p} : \|\boldsymbol{X}\|_F = 1\}$

For a detailed introduction on optimization on Riemannian manifolds: see [2].

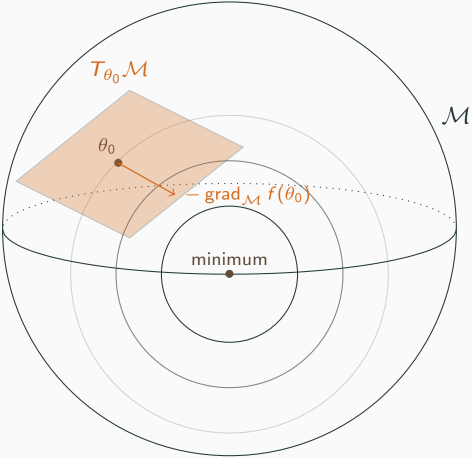
Optimization on a manifold



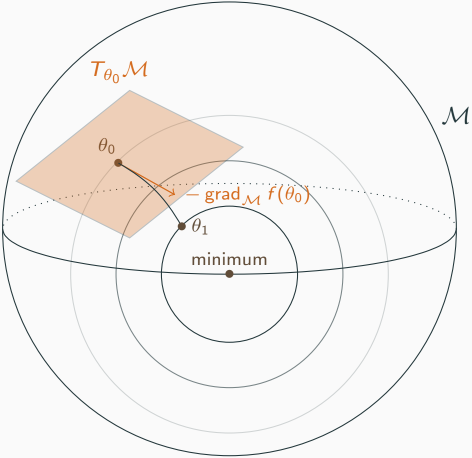
Optimization on a manifold



Optimization on a manifold



Optimization on a manifold



Existing work (1/2)

$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{R}^p$ realizations of $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, $\Sigma \in \mathcal{S}_p^{++}$ (set of $p \times p$ symmetric positive definite matrices).

Step 2: maximum likelihood estimator

$$\theta = \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T . \tag{1}$$

Step 3: Riemannian geometry of centered Gaussian distributions

$\mathcal{S}_p^{++}$ with the Fisher information metric:

$\forall \boldsymbol{\xi}_\Sigma, \boldsymbol{\eta}_\Sigma$ in the tangent space at $\Sigma \in \mathcal{S}_p^{++}$

$$\langle \boldsymbol{\xi}_\Sigma, \boldsymbol{\eta}_\Sigma \rangle_\Sigma^{\text{FIM}} = \text{Tr}\left(\Sigma^{-1} \boldsymbol{\xi}_\Sigma \Sigma^{-1} \boldsymbol{\eta}_\Sigma \right) . \tag{2}$$

NB: invariance property by affine transformations:

$$\langle \mathrm{D}\, \phi(\Sigma)[\boldsymbol{\xi}_\Sigma], \mathrm{D}\, \phi(\Sigma)[\boldsymbol{\eta}_\Sigma] \rangle_{\phi(\Sigma)}^{\text{FIM}} = \langle \boldsymbol{\xi}_\Sigma, \boldsymbol{\eta}_\Sigma \rangle_\Sigma^{\text{FIM}} . \tag{3}$$

where $\phi(\Sigma) = \boldsymbol{A} \Sigma \boldsymbol{A}^T$, $\forall \boldsymbol{A} \in \mathbb{R}^{p \times p}$ invertible.

Existing work (2/2)

Step 3

Riemannian distance between Σ_1 and Σ_2 in $\mathcal{S}_p^{++}$:

$$d_{\mathcal{S}_p^{++}}(\Sigma_1, \Sigma_2) = \left\| \log \left(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}} \right) \right\|_2 . \tag{4}$$

NB: invariance property by affine transformations:

$$d_{\mathcal{S}_p^{++}}(\phi(\Sigma_1), \phi(\Sigma_2)) = d_{\mathcal{S}_p^{++}}(\Sigma_1, \Sigma_2) \tag{5}$$

Riemannian mean of a set $\{\Sigma_i\}$:

$$\Sigma_{\text{mean}} = \underset{\Sigma \in \mathcal{S}_p^{++}}{\arg\min} \sum_i d^2_{\mathcal{S}_p^{++}}(\Sigma, \Sigma_i). \tag{6}$$

Enough to apply a *K-means* or a *Nearest centroïd classifier*.
For a full description of the manifold $\mathcal{S}_p^{++}$ and its associated center of mass: see [3], [4].

Geodesic triangles for classification problems

Geodesic triangles for machine learning

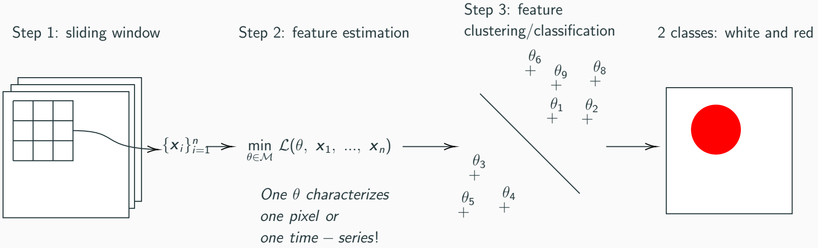


Figure 6: Clustering/classification pipeline.

Statistical model

Let $x_1, \cdots, x_n \in \mathbb{R}^p$ distributed as $x \sim \mathcal{N}(\mu, \Sigma)$ with $\mu \in \mathbb{R}^p$, $\Sigma \in \mathcal{S}_p^{++}$.
Goal: classify $\theta = (\mu, \Sigma)$.

Riemannian geometry of Gaussian distributions

Riemannian geometry of non-centered Gaussian distributions

$\mathbb{R}^p \times \mathcal{S}_p^{++}$ with the Fisher information metric: $\forall \xi = (\boldsymbol{\xi}_{\boldsymbol{\mu}}, \boldsymbol{\xi}_\Sigma), \eta = (\boldsymbol{\eta}_{\boldsymbol{\mu}}, \boldsymbol{\eta}_\Sigma)$ in the tangent space

$$\langle \xi, \eta \rangle_{(\boldsymbol{\mu},\Sigma)}^{\text{FIM}} = \boldsymbol{\xi}_{\boldsymbol{\mu}}^T \Sigma^{-1} \boldsymbol{\eta}_{\boldsymbol{\mu}} + \frac{1}{2} \operatorname{Tr}\left(\Sigma^{-1} \boldsymbol{\xi}_\Sigma \Sigma^{-1} \boldsymbol{\eta}_\Sigma\right). \tag{7}$$

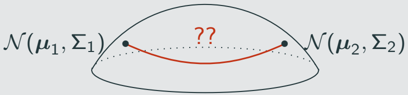
NB: invariance property by affine transformations:

$$\langle \mathrm{D}\,\phi(\boldsymbol{\mu},\Sigma)[\xi], \mathrm{D}\,\phi(\boldsymbol{\mu},\Sigma)[\eta] \rangle_{(\boldsymbol{\mu},\Sigma)}^{\text{FIM}} = \langle \xi, \eta \rangle_{(\boldsymbol{\mu},\Sigma)}^{\text{FIM}} \tag{8}$$

with $\phi(\boldsymbol{\mu},\Sigma) = (\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{\mu}_0, \boldsymbol{A}\Sigma\boldsymbol{A}^T)$, $\forall \boldsymbol{A} \in \mathbb{R}^{p\times p}$ invertible, $\forall \boldsymbol{\mu}_0 \in \mathbb{R}^p$.

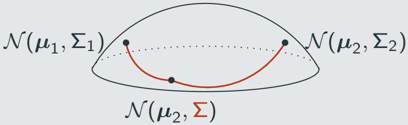
Problem

Problem: this Riemannian geometry is not fully known... (see [5], [6])



Geodesic triangles for machine learning

Solution: use of geodesic triangles



Divergence δ: arc length of the path between $(\boldsymbol{\mu}_1, \Sigma_1)$ and $(\boldsymbol{\mu}_2, \Sigma_2)$.

$$\delta_c : \quad (\boldsymbol{\mu}_1, \Sigma_1) \to (\boldsymbol{\mu}_1, c\Sigma_1) \to (\boldsymbol{\mu}_2, \Sigma_2), \qquad \forall c > 0$$

$$\delta_\perp : \quad (\boldsymbol{\mu}_1, \Sigma_1) \to (\boldsymbol{\mu}_1, \Sigma_1 + \Delta\boldsymbol{\mu}\Delta\boldsymbol{\mu}^T) \to (\boldsymbol{\mu}_2, \Sigma_2), \quad \Delta\boldsymbol{\mu} = \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$$

NB: both divergences are invariant by affine transformations.

Riemannian center of mass $(\boldsymbol{\mu}_{\mathbf{mean}}, \Sigma_{\mathbf{mean}})$ of $\{(\boldsymbol{\mu}_i, \Sigma_i)\}$

$$(\boldsymbol{\mu}_{\text{mean}}, \Sigma_{\text{mean}}) = \underset{(\boldsymbol{\mu},\Sigma)\in\mathbb{R}^p\times\mathcal{S}_p^{++}}{\arg\min} \sum_i \delta^2\left((\boldsymbol{\mu}, \Sigma), (\boldsymbol{\mu}_i, \Sigma_i)\right) \tag{9}$$

Riemannian optimization

Minimize $f : (\boldsymbol{\mu}, \Sigma) \to \mathbb{R}$.

Proposition (Riemannian gradient)

The Riemannian gradient of f at $(\boldsymbol{\mu}, \Sigma)$ is

$$\operatorname{grad} f(\boldsymbol{\mu}, \Sigma) = \left(\Sigma \boldsymbol{G}_{\boldsymbol{\mu}}, \Sigma\left(\boldsymbol{G}_\Sigma + \boldsymbol{G}_\Sigma^T\right)\Sigma\right)$$

where $\operatorname{grad}_\epsilon f(\boldsymbol{\mu}, \Sigma) = (\boldsymbol{G}_{\boldsymbol{\mu}}, \boldsymbol{G}_\Sigma) \in \mathbb{R}^p \times \mathbb{R}^{p\times p}$ *is the Euclidean gradient of f.*

Proposition (Second order retraction)

A second order retraction at $(\boldsymbol{\mu}, \Sigma)$ of ξ in the tangent space is

$$R_{(\boldsymbol{\mu},\Sigma)}(\boldsymbol{\xi}_{\boldsymbol{\mu}}, \boldsymbol{\xi}_\Sigma) = \left(\boldsymbol{\mu} + \boldsymbol{\xi}_{\boldsymbol{\mu}} + \frac{1}{2}\boldsymbol{\xi}_\Sigma \Sigma^{-1}\boldsymbol{\xi}_{\boldsymbol{\mu}}, \Sigma + \boldsymbol{\xi}_\Sigma + \frac{1}{2}\left[\boldsymbol{\xi}_\Sigma \Sigma^{-1}\boldsymbol{\xi}_\Sigma - \boldsymbol{\xi}_{\boldsymbol{\mu}}\boldsymbol{\xi}_{\boldsymbol{\mu}}^T\right]\right).$$

Riemannian optimization

Riemannian gradient descent

Input : Initial iterate $(\boldsymbol{\mu}_1, \Sigma_1)$.
Output: Sequence of iterates $\{(\boldsymbol{\mu}_k, \Sigma_k)\}$.
$k := 1$;
while *no convergence* **do**
  Compute a step size α and set
  $(\boldsymbol{\mu}_{k+1}, \Sigma_{k+1}) := R_{(\boldsymbol{\mu}_k, \Sigma_k)}(-\alpha \operatorname{grad} f(\boldsymbol{\mu}_k, \Sigma_k))$;
  $k := k + 1$;
end

Algorithm 1: Riemannian gradient descent

Breizhcrops dataset [1]:

- more than 600 000 crop time series across the whole Brittany,
- 9 classes,
- 13 spectral bands.

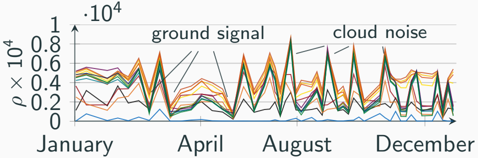


Figure 7: Reflectances of a Sentinel-2 time series.

Estimator	Geometry	OA (%)	AA (%)
$\boldsymbol{X}$	$\mathbb{R}^{p \times n}$	10.1	18.5
Mean $\hat{\boldsymbol{\mu}}$	$\mathbb{R}^{p}$	13.2	14.8
Covariance matrix $\hat{\Sigma}$	$\mathcal{S}_p^{++}$	43.9	28.1
Centered covariance matrix $\hat{\Sigma}$	$\mathcal{S}_p^{++}$	46.7	30.1
Proposed - $(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$	δ_c	**54.3**	**37.0**
Proposed - $(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$	$\delta_\perp$	53.3	35.7

Table 1: Accuracies of *Nearest centroïd classifiers* on the *Breizhcrops* dataset.
OA = Overall Accuracy, AA = Average Accuracy

Conclusion

Conclusion

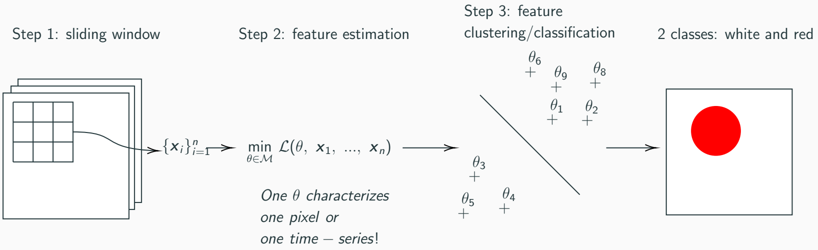


Figure 8: Clustering/classification pipeline.

Theoretical contributions:

- new divergences: δ_c, $\delta_\perp$
- new algorithm to compute Riemannian centers of mass.

Application on a real dataset of multispectral time-series classification: *Breizhcrops*.

ON THE USE OF GEODESIC TRIANGLES BETWEEN GAUSSIAN DISTRIBUTIONS FOR CLASSIFICATION PROBLEMS

IEEE ICASSP 2022 - Singapore

A. Collas[1], F. Bouchard[2], G. Ginolhac[3], A. Breloy[4], C. Ren[1], J.-P. Ovarlez[1,5]

[1]SONDRA, CentraleSupélec, Univ. Paris-Saclay
[2]CNRS, L2S, CentraleSupélec, Univ. Paris-Saclay
[3]LISTIC, Univ. Savoie Mont Blanc
[4]LEME, Univ. Paris Nanterre
[5]DEMR, ONERA, Univ. Paris-Saclay

E-mail: antoine.collas@centralesupelec.fr

References i

M. Rußwurm, C. Pelletier, M. Zollner, S. Lefèvre, and M. Körner, "Breizhcrops: A time series dataset for crop type mapping," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences ISPRS (2020)*, 2020.

P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton University Press, 2008.

L. T. Skovgaard, "A Riemannian geometry of the multivariate Normal model," *Scandinavian Journal of Statistics*, vol. 11, no. 4, pp. 211–223, 1984, ISSN: 03036898, 14679469. [Online]. Available: http://www.jstor.org/stable/4615960.

M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 3, pp. 735–747, 2005. DOI: `10.1137/S0895479803436937`. eprint: `https://doi.org/10.1137/S0895479803436937`. [Online]. Available: `https://doi.org/10.1137/S0895479803436937`.

M. Calvo and J. M. Oller, "An explicit solution of information geodesic equations for the multivariate normal model," *Statistics & Risk Modeling*, vol. 9, no. 1-2, pp. 119–138, 1991. DOI: `doi:10.1524/strm.1991.9.12.119`. [Online]. Available: `https://doi.org/10.1524/strm.1991.9.12.119`.

References iii

M. Tang, Y. Rong, J. Zhou, and X. Li, "Information geometric approach to multisensor estimation fusion," *IEEE Transactions on Signal Processing*, vol. 67, no. 2, pp. 279–292, 2019. DOI: 10.1109/TSP.2018.2879035.