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

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

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

Time series for remote sensing and classification

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

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

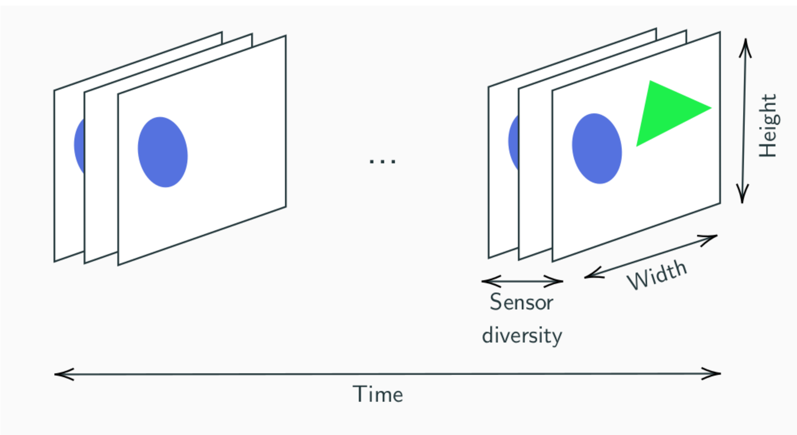

Figure 1. Multivariate image time series.

Applications: disaster assessment, activity monitoring, land cover mapping, crop type mapping, ...

Classification pipeline

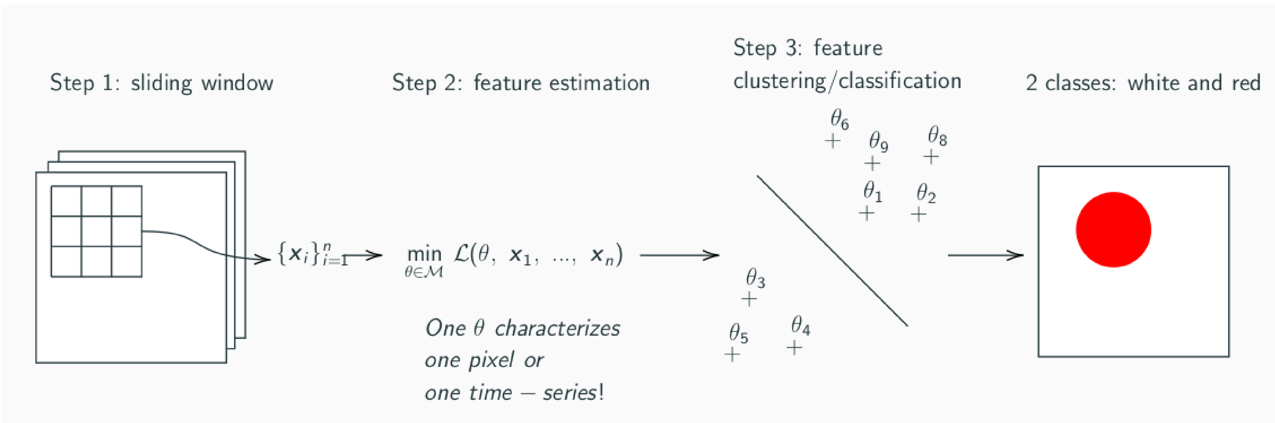

Figure 2. Classification pipeline.

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

Existing work and Riemannian geometry

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

Step 2: maximum likelihood estimator:

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

Step 3: Riemannian manifold of centered Gaussian distributions:

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

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

- Riemannian distance

$$d_{\mathcal{S}_p^{++}}(\boldsymbol{\Sigma}_l, \boldsymbol{\Sigma}_m) = \left\| \log\left(\boldsymbol{\Sigma}_l^{-\frac{1}{2}} \boldsymbol{\Sigma}_m \boldsymbol{\Sigma}_l^{-\frac{1}{2}} \right) \right\|_2. \tag{3}$$

- Riemannian center of mass of a set $\{\boldsymbol{\Sigma}_i\}$

$$\boldsymbol{\Sigma}_{\text{mean}} = \underset{\boldsymbol{\Sigma} \in \mathcal{S}_p^{++}}{\arg\min} \sum_i d^2_{\mathcal{S}_p^{++}}(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}_i). \tag{4}$$

For a full description of the manifold $\mathcal{S}_p^{++}$ and its associated center of mass: see [1, 2].

The Riemannian manifold of non-centered Gaussian distributions

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

$$\langle \xi, \eta \rangle_{(\boldsymbol{\mu}, \boldsymbol{\Sigma})}^{\text{FIM}} = \boldsymbol{\xi}_{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\eta}_{\boldsymbol{\mu}} + \frac{1}{2} \operatorname{Tr}\left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}_{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\eta}_{\boldsymbol{\Sigma}}\right). \tag{5}$$

Problem: this Riemannian geometry is not fully known... (see [3, 4])

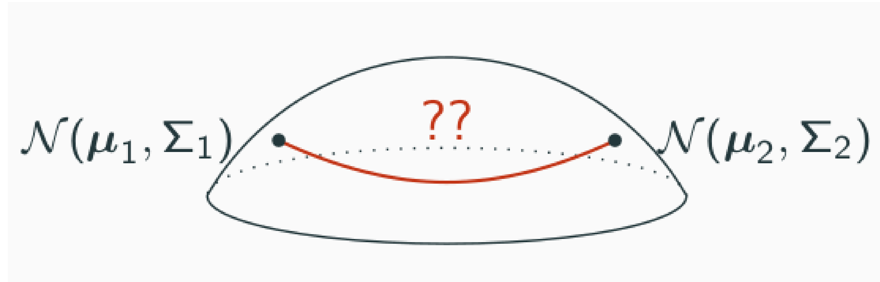

Figure 3. The geodesic between two non-centered Gaussian distributions is unknown in general.

Geodesic triangles for classification problems

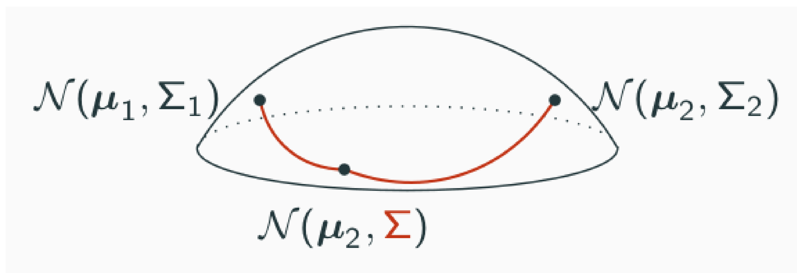

Figure 4. A geodesic triangle.

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

$$\begin{aligned} \delta_c :&\quad (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \to (\boldsymbol{\mu}_1, c\boldsymbol{\Sigma}_1) \to (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), && \forall c > 0 \\ \delta_\perp :&\quad (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \to (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1 + \Delta\boldsymbol{\mu}\Delta\boldsymbol{\mu}^T) \to (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), && \Delta\boldsymbol{\mu} = \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1 \end{aligned}$$

Center of mass and Riemannian optimization

Riemannian center of mass $(\boldsymbol{\mu}_{\text{mean}}, \boldsymbol{\Sigma}_{\text{mean}})$ of a set $\{(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\}$

$$(\boldsymbol{\mu}_{\text{mean}}, \boldsymbol{\Sigma}_{\text{mean}}) = \underset{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathbb{R}^p \times \mathcal{S}_p^{++}}{\arg\min} \sum_i \delta^2\left((\boldsymbol{\mu}, \boldsymbol{\Sigma}), (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\right) \tag{6}$$

Algorithm to minimize a real-valued function f defined on $\mathbb{R}^p \times \mathcal{S}_p^{++}$:

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

Algorithm 1: Riemannian gradient descent

- $\operatorname{grad} f(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the Riemannian gradient of f at $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ computed in Proposition 1,
- $R_{(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}$ is a second order retraction at $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ derived in Proposition 2.

For a detailed introduction to optimization on Riemannian manifolds: see [5].

Application

Breizhcrops dataset [6]:

- more than 600 000 crop time series across the whole Brittany taken by the Sentinel-2 satellite,
- 9 classes: barley, wheat, rapeseed, corn, sunflower, orchards, nuts, permanent meadows and temporary meadows,
- 13 spectral bands.

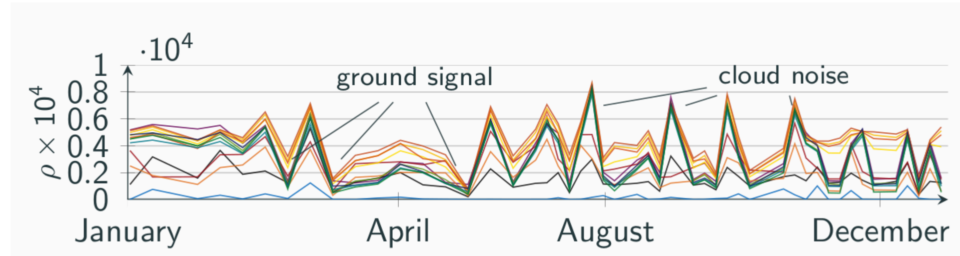

Figure 5. Reflectances of a Sentinel-2 time series from the *Breizhcrops* dataset.

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

Table 1. Accuracies of *Nearest centroïd classifiers* on the *Breizhcrops* dataset.

We denote the columns of a time-series by $\boldsymbol{X}_j = [[\boldsymbol{X}_j]_{:,1}, \cdots, [\boldsymbol{X}_j]_{:,n}] \in \mathbb{R}^{p\times n}$. Different estimators/geometries are considered:

- $\boldsymbol{X}_j$: raw time-series with the Euclidean distance $d(\boldsymbol{X}_l, \boldsymbol{X}_m) = \|\boldsymbol{X}_l - \boldsymbol{X}_m\|_F$ and the arithmetic mean $\boldsymbol{X}_{\text{mean}} = \frac{1}{M}\sum_{j=1}^{M} \boldsymbol{X}_j$,
- $\hat{\boldsymbol{\mu}}_j = \frac{1}{n}\sum_{i=1}^{n} [\boldsymbol{X}_j]_{:,i}$: temporal mean with the Euclidean distance $d(\hat{\boldsymbol{\mu}}_l, \hat{\boldsymbol{\mu}}_m) = \|\hat{\boldsymbol{\mu}}_l - \hat{\boldsymbol{\mu}}_m\|_2$ and the arithmetic mean $\hat{\boldsymbol{\mu}}_{\text{mean}} = \frac{1}{M}\sum_{j=1}^{M} \hat{\boldsymbol{\mu}}_j$,
- $\hat{\boldsymbol{\Sigma}}_j = \frac{1}{n}\sum_{i=1}^{n} [\boldsymbol{X}_j]_{:,i} [\boldsymbol{X}_j]_{:,i}^T$: temporal covariance matrix with the distance (3) and its associated Riemannian mean (4),
- $\hat{\boldsymbol{\Sigma}}_j = \frac{1}{n}\sum_{i=1}^{n} \left([\boldsymbol{X}_j]_{:,i} - \hat{\boldsymbol{\mu}}_j\right)\left([\boldsymbol{X}_j]_{:,i} - \hat{\boldsymbol{\mu}}_j\right)^T$: temporal centered covariance matrix with the distance (3) and its associated Riemannian mean (4),
- $(\hat{\boldsymbol{\mu}}_j, \hat{\boldsymbol{\Sigma}}_j)$: temporal mean and centered covariance matrix with the divergence δ_c and its associated Riemannian center of mass (6),
- $(\hat{\boldsymbol{\mu}}_j, \hat{\boldsymbol{\Sigma}}_j)$: temporal mean and centered covariance matrix with the divergence $\delta_\perp$ and its associated Riemannian center of mass (6).

References

[1] L. T. Skovgaard, "A Riemannian geometry of the multivariate Normal model," *Scandinavian Journal of Statistics*, vol. 11, no. 4, pp. 211–223, 1984.

[2] 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.

[3] 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.

[4] 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.

[5] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. 2008.

[6] 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.

antoine.collas@centralesupelec.fr