Robust Clustering for Satellite Images Time-Series

Antoine Collas - 3ème année - SONDRA - ED STIC Paris-Saclay

17 janvier 2022

Supervisors:
Jean-Philippe Ovarlez - Directeur de thèse - ONERA/SONDRA
Guillaume Ginolhac - Directeur de thèse - LISTIC
Chengfang Ren - Co-encadrant - SONDRA
Arnaud Breloy - Co-encadrant - LEME

Context

Context of the PhD

In the last few years many images have been taken from the earth with different technologies (SAR, multi-spectral/hyperspectral imaging, ...).

Problematics

The objective is to extract **semantic information** in these new data. More particularly we focus on 2 specific topics:

- semantic segmentation (spatial information),
- time-series clustering/classification (temporal information).

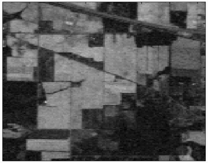


Figure 1: Raw image.

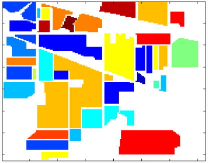


Figure 2: Segmented image, one color = one class (grass, woods, ...).

Example of multi-spectral time series

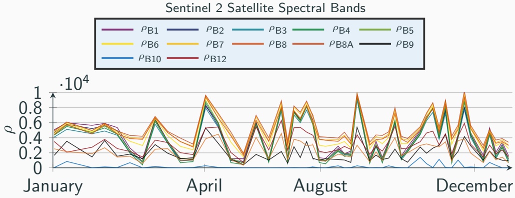


Figure 3: Example of reflectances ρ of a Sentinel-2 time series of **meadows** from *Breizhcrops* dataset.

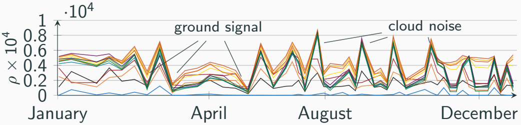


Figure 4: Example of reflectances ρ of a Sentinel-2 time series of **corn** from *Breizhcrops* dataset.

Clustering/classification pipeline and Riemannian geometry

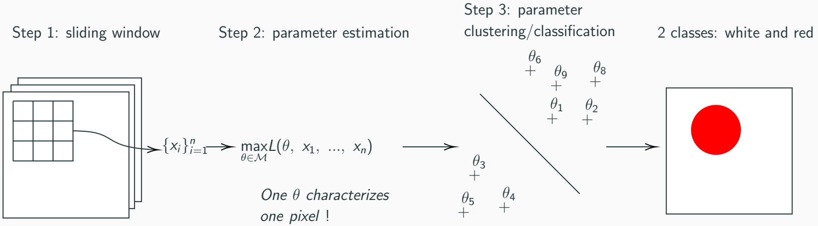


Figure 5: Clustering/classification pipeline on an image.

Examples of θ:

$\theta = \Sigma$ a covariance matrix, $\theta = (\boldsymbol{\mu}, \Sigma)$ a vector and a covariance matrix, $\theta = (\tau_i, \boldsymbol{U})$ a scalar and an orthogonal matrix...

Clustering/classification pipeline and Riemannian geometry

Clustering/classification and Riemannian geometry

The statistical model depends on $\theta \in \mathcal{M}$, a *Riemannian manifold*:

- step 2: maximization of the likelihood L over $\mathcal{M}$,
- step 3: computing distances and centers of mass on $\mathcal{M}$.

Existing work

Let $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{C}^p$ be data points distributed as $\boldsymbol{x} \sim \mathbb{C}\mathcal{N}(0, \Sigma)$.

Step 2: maximum likelihood estimator:

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

Step 3: Riemannian distance on $\mathcal{H}_p^{++}$ (geodesic distance):

$$d(\Sigma_1, \Sigma_2) = \left\| \log \left(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}} \right) \right\|_2. \tag{2}$$

Riemannian geometry and objectives

What is a Riemannian manifold ?

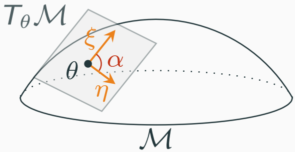


Figure 6: A Riemannian manifold.

Curvature induced by:

- constraints, *e.g.* the sphere: $\|\boldsymbol{x}\| = 1$,
- the Riemannian metric, *e.g.* on $\mathcal{H}_p^{++}$: $\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle_\Sigma^\mathcal{M} = \mathrm{Tr}(\Sigma^{-1}\boldsymbol{\xi}\Sigma^{-1}\boldsymbol{\eta})$.

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

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

Step 2: objectives for parameter estimation

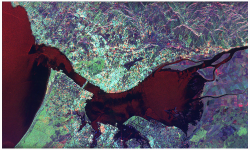


Figure 7: Example of a SAR image (from `nasa.gov`).

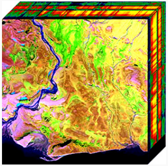


Figure 8: Example of a hyperspectral image (from `nasa.gov`).

Objectives:

- Develop **robust estimators**, *i.e.* estimators that work well with non Gaussian data because of the high resolution of images.
- Develop **regularized estimators**, *i.e.* estimators that handle high dimension of hyperspectral images.

Step 3: objectives for parameter clustering/classification

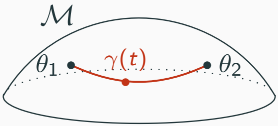


Figure 9: Distance:
length of the geodesic γ.

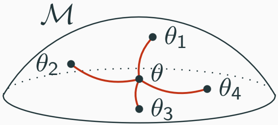


Figure 10: Center of mass:
$\theta = \arg\min_{\theta \in \mathcal{M}} \sum_i d_\gamma^2(\theta, \theta_i)$.

Objectives:

Develop distances

- that **respect the geometry** of $\mathcal{M}$ (*e.g.* orthogonality constraints),
- and that are **related to the chosen statistical distributions**.

Study of a ”low rank” model

Study of a "low rank" model [4]

Let $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{C}^p$ be data points.

Assumption

The signal belongs to a $k < p$ dimensional subspace denoted span($\boldsymbol{U}$).

Model

$$\underbrace{\boldsymbol{x}_i}_{\in \mathbb{C}^p} | \tau_i \stackrel{d}{=} \underbrace{\sqrt{\tau_i}\, \boldsymbol{U} \boldsymbol{g}_i}_{\text{signal} \in \text{span}(\boldsymbol{U})} + \underbrace{\boldsymbol{n}_i}_{\text{noise} \in \mathbb{C}^p} \tag{3}$$

where $\boldsymbol{g}_i \sim \mathbb{C}\mathcal{N}(0, \boldsymbol{I}_k)$ and $\boldsymbol{n}_i \sim \mathbb{C}\mathcal{N}(0, \boldsymbol{I}_p)$ are independent;
$\boldsymbol{\tau} \in (\mathbb{R}_*^+)^n$ contains the unknown deterministic textures τ_i; and
$\boldsymbol{U} \in \mathbb{C}^{p \times k}$ is an orthogonal basis of the subspace.

MLE and intrinsic Cramèr-Rao bound [4]

Maximum likelihood estimation (MLE)

Maximization of the likelihood while respecting the constraints:

- $\boldsymbol{U}$: orthogonal basis of the subspace (and thus invariant by rotation !)
- $\boldsymbol{\tau} \in (\mathbb{R}_*^+)^n$ (positivity constraint)

$$\max_{\boldsymbol{U}, \boldsymbol{\tau} \in \mathcal{M}} L(\boldsymbol{U}, \boldsymbol{\tau}) \tag{4}$$

We proposed a Riemannian stochastic gradient descent for this problem.

Bounds

Study of the performance through intrinsic Cramèr-Rao bounds:

$$\overbrace{\mathbb{E}[d^2_{\mathrm{Gr}_{p,k}}(\mathrm{span}(\hat{\boldsymbol{U}}), \mathrm{span}(\boldsymbol{U}))]}^{\text{subspace estimation error}} \geq \frac{(p-k)k}{nc_{\boldsymbol{\tau}}} \approx \frac{(p-k)k}{n \times \mathrm{SNR}} \tag{5}$$

$$\underbrace{\mathbb{E}[d^2_{(\mathbb{R}_*^+)^n}(\hat{\boldsymbol{\tau}}, \boldsymbol{\tau})]}_{\text{texture estimation error}} \geq \frac{1}{k} \sum_{i=1}^{n} \frac{(1+\tau_i)^2}{\tau_i^2} \tag{6}$$

Illustration [4]

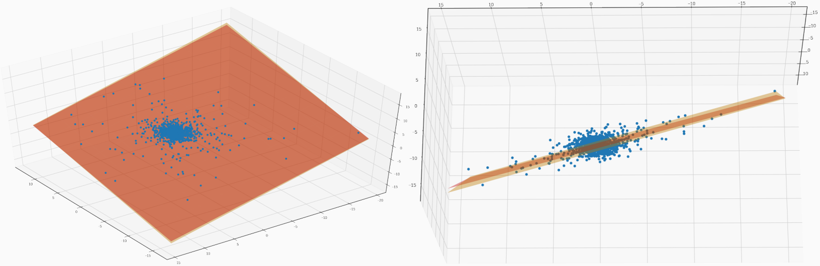


Figure 11: Scatter plot of samples $\{\boldsymbol{x}_i\}_{i=1}^{1000}$ with real and estimated subspaces respectively in orange and red in the case $\mathbb{E}[\tau_i] = 10$.

Remark

Both subspaces are really close !

K-means++ on a Riemannian manifold [4]

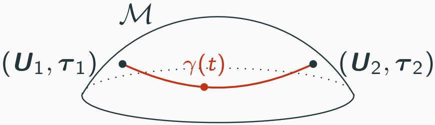


Figure 12: Distance.

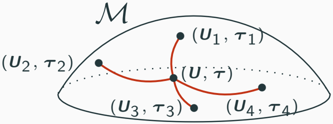


Figure 13: Center of mass $(\boldsymbol{U}, \boldsymbol{\tau})$.

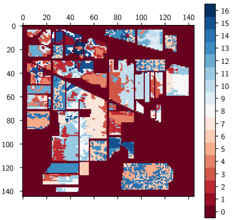


Figure 14: Euclidean *K-means++*: OA $= 31.2\%$.

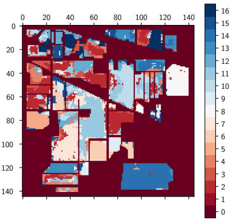


Figure 15: Proposed *K-means++*: OA $= 47.2\%$

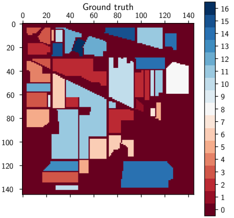


Figure 16: Ground truth.

A Tyler-type estimator

A Tyler-type estimator using manifold optimization [1]

Problem

Let $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{C}^p$ distributed as

$$\boldsymbol{x}_i \sim \mathbb{C}\mathcal{N}(\boldsymbol{\mu}, \tau_i \Sigma) \tag{7}$$

with $\boldsymbol{\mu} \in \mathbb{C}^p$, $\boldsymbol{\tau} \in (\mathbb{R}_*^+)^n$ and $\Sigma \in \mathcal{H}_p^{++} (\Sigma \succ 0)$.

- $\boldsymbol{\mu} = 0$: Tyler's estimator converges to the maximum likelihood estimator.
- $\boldsymbol{\mu}$ unknown: no estimator realizing the maximum likelihood estimator exists ...

Solution: maximizing the likelihood on a Riemannian manifold

$$\max_{\boldsymbol{\mu} \in \mathbb{C}^p, \Sigma \succ 0, \tau_i > 0} L(\boldsymbol{\mu}, \Sigma, \boldsymbol{\tau}) \tag{8}$$

We proposed a Riemannian conjugate gradient for this problem.

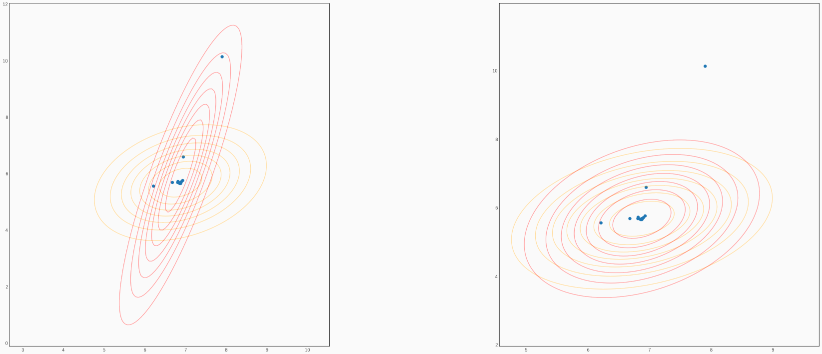


Figure 17: Scatter plot of samples $\{\boldsymbol{x}_i\}_{i=1}^{10}$ with real and estimated p.d.f respectively in orange and red. Left are the Gaussian estimators. Right are our estimators.

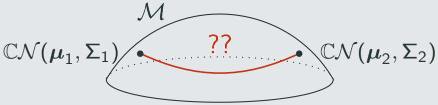
Geodesic triangles for machine learning

Geodesic triangles for machine learning [2]

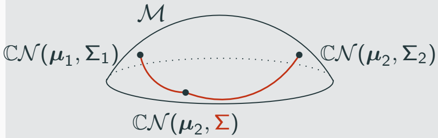
Let $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{C}^p$ distributed as $\boldsymbol{x} \sim \mathbb{C}\mathcal{N}(\boldsymbol{\mu}, \Sigma)$.

Problem

The Riemannian geometry of the Gaussian distributions is not fully known ...



Solution: use of geodesic triangles



With a cleverly chosen $\boldsymbol{\Sigma}$!
Induces a divergence and a Riemannian center of mass !

Geodesic triangles for machine learning [2]

Breizhcrops dataset:

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

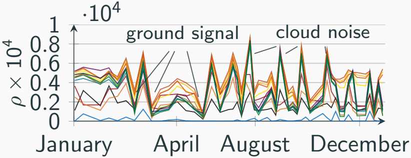


Figure 18: Example of reflectances ρ of a Sentinel-2 time series.

Estimator of $\boldsymbol{X}$	Geometry	Overall accuracy (%)
$\boldsymbol{X}$	$\mathbb{R}^{p \times n}$	10.1
$\hat{\boldsymbol{\mu}}$	$\mathbb{R}^{p}$	13.2
$\hat{\Sigma}$, ($\boldsymbol{\mu}$ known)	$\mathcal{S}_p^{++}$	43.9
$\hat{\Sigma}$, ($\boldsymbol{\mu}$ unknown)	$\mathcal{S}_p^{++}$	46.7
(proposed) $(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$	$\mathcal{N}^p$ with $\delta^2_{c,\mathcal{N}^p}$	**54.3**
(proposed) $(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$	$\mathcal{N}^p$ with $\delta^2_{\perp,\mathcal{N}^p}$	**53.3**

Table 1: Results of *Nearest centroïd* algorithms with different estimators and geometries on the *Breizhcrops* dataset.

Conclusion

Conclusion

Theoretical contributions on the whole clustering/classification pipeline that go beyond the Gaussian assumption with known location:

- step 2: new estimators: "low rank" estimator [4], Tyler-type estimator [1]
- step 2 analysis: new intrinsic Cramér-Rao bounds: "low rank" model [4],
- step 3: new Riemannian distances and center of mass: "low rank" model [4], geodesic triangles [2]

Applications on datasets of earth observation:

- unsupervised learning: Riemannian *K-means++* on *Indian pines* hyperspectral image,
- supervised learning: Riemannian *Nearest centroid* on *Breizhcrops* multispectral times series.

Perspectives

2 perspectives:

- on the Tyler-type estimator:
  - accelerate the estimation: 2nd order optimizer, ...
  - define a distance/divergence to perform machine learning,
  - apply this statistical model on a dataset,
- metric learning: instead of defining a metric a-priori, learning the metric from data.

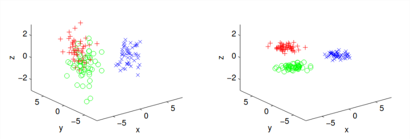


Figure 19: Example of metric learning.

Conferences

- **A. Collas**, F. Bouchard, A. Breloy, C. Ren, G. Ginolhac, and J.-P. Ovarlez, "A tyler-type estimator of location and scatter leveraging riemannian optimization,", ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). DOI: 10.1109/ICASSP39728.2021.9414974.
- **A. Collas**, F. Bouchard, G. Ginolhac, A. Breloy, C. Ren, and J.-P. Ovarlez, "On the use of geodesic triangles between gaussian distributions for classification problems,", Submitted to ICASSP 2022, minor review - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Journals

- A. Mian, **A. Collas**, A. Breloy, G. Ginolhac, and J.-P. Ovarlez, "Robust low-rank change detection for multivariate sar image time series," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3545–3556, 2020. DOI: 10.1109/JSTARS.2020.2999615.
- **A. Collas**, F. Bouchard, A. Breloy, G. Ginolhac, C. Ren, and J.-P. Ovarlez, "Probabilistic pca from heteroscedastic signals: Geometric framework and application to clustering," *IEEE Transactions on Signal Processing*, vol. 69, pp. 6546–6560, 2021. DOI: 10.1109/TSP.2021.3130997.

Ongoing

A. Collas, A. Breloy, C. Ren, G. Ginolhac, and J.-P. Ovarlez, *Riemannian classification approach to non-centered mixture of scaled gaussian*,

Software and workshops

Software:

- contribution to *pyManopt*[1]: a Python toolbox for optimization on Riemannian manifolds ,
- creation of *pyCovariance*: a Python toolbox for covariance estimation and clustering/classification using Riemannian geometry.

Workshops:

- talk at "Statistical Learning for Signal and Image Processing (SLSIP) Workshop", A German-Finnish-French Workshop, Rüdesheim October 2021,
- project of Riemannian optimization for optimal transport at "LOGML 2021", mentored by Bamdev Mishra, creator of *Manopt* the leading toolbox of optimization of Riemannian manifolds.

[1] `https://github.com/pymanopt/pymanopt`

Appendices

Numerical experiment [4]

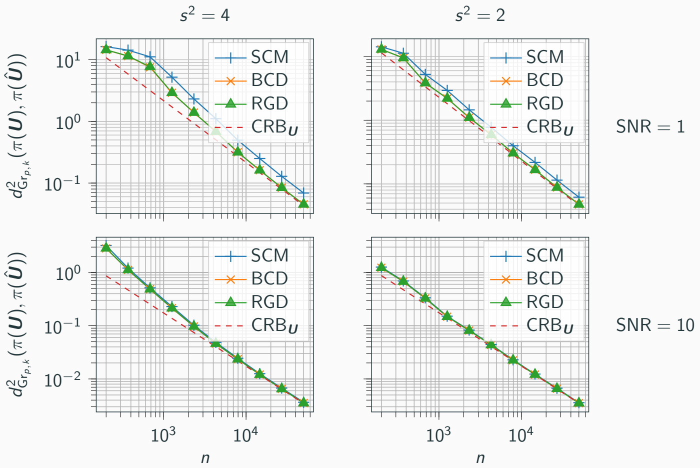


Figure 20: MSE over $N = 100$ simulated sets $\{\boldsymbol{x}_i\}$ ($p = 100$ and $k = 20$) with respect to the number of samples n for the considered estimators. s^2 controls the heterogeneity of the textures.

Numerical experiment [1]

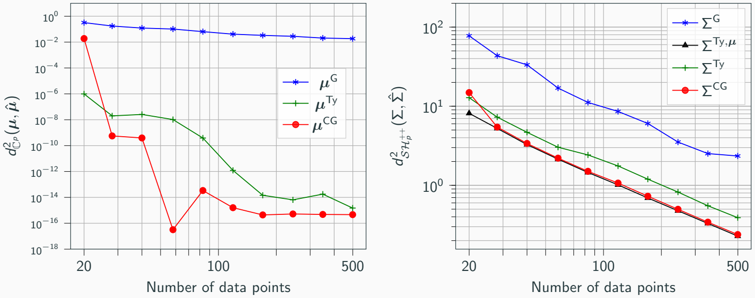


Figure 21: Subspace estimation error over $N = 200$ simulated sets $\{\boldsymbol{x}_i\}$ ($p = 10$) with respect to the number of samples n for different estimators. CG is the Riemannian Conjugate Gradient and "Ty,$\boldsymbol{\mu}$" is the Tyler's estimator with $\boldsymbol{\mu}$ known.