Optimization and statistical learning using Riemannian geometry and application to remote sensing

Antoine Collas
PhD begun in October 2019, defense expected in September 2022
SONDRA lab, CentraleSupélec, Univ. Paris-Saclay

Supervisors:
Jean-Philippe Ovarlez - Director - ONERA/SONDRA
Guillaume Ginolhac - Director - LISTIC
Chengfang Ren - SONDRA
Arnaud Breloy - LEME

Table of contents

Context

Context of the PhD

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...).

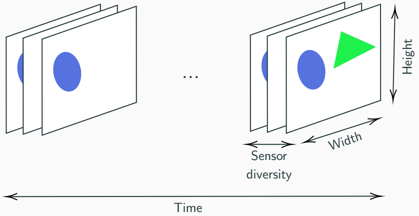


Figure 1: Multivariate image time series.

Applications

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

Example of a hyperspectral image

Indian pines dataset:

- 145×145 pixels, 200 spectral bands,
- 16 classes (corn, grass, wood, ...).

Figure 2: Raw image.

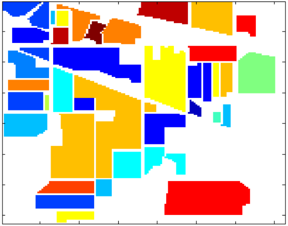


Figure 3: Segmented image, one color = one class.

Example of multi-spectral time series

Breizhcrops dataset[1]:

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

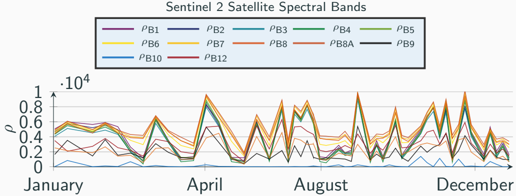


Figure 4: Reflectances ρ of a time series of **meadows**.

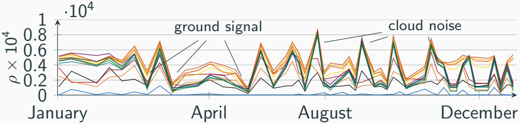


Figure 5: Reflectances ρ of a time series of **corn**.

[1] `https://breizhcrops.org/`

Clustering/classification pipeline and Riemannian geometry

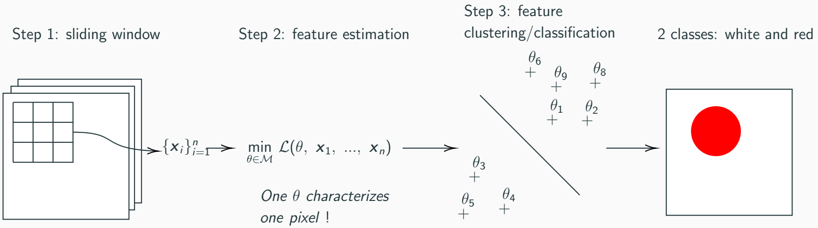


Figure 6: Clustering/classification pipeline.

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

$\theta \in \mathcal{M}$, a *Riemannian manifold* (contraints and non-constant metric):

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

Existing work (e.g. in BCI classification)

$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{R}^p$ realizations of $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, $\Sigma \in \mathcal{S}_p^{++}$.
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 distance on $\mathcal{S}_p^{++}$ (geodesic distance):

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

Riemannian geometry and objectives

What is a Riemannian manifold ?

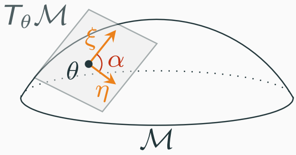


Figure 7: A Riemannian manifold.

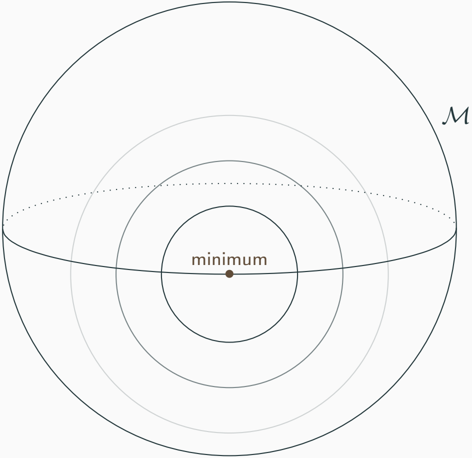
Curvature induced by:

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

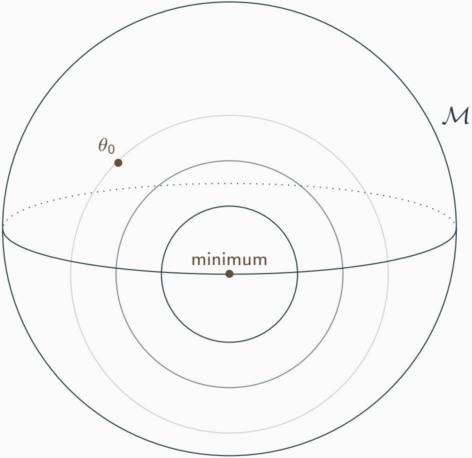
Examples of Riemannian manifolds $\mathcal{M}$:

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

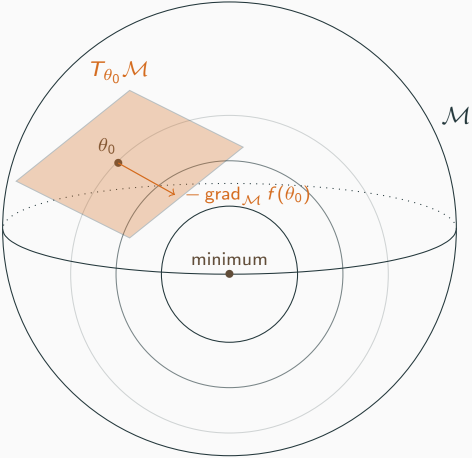
Optimization on a manifold



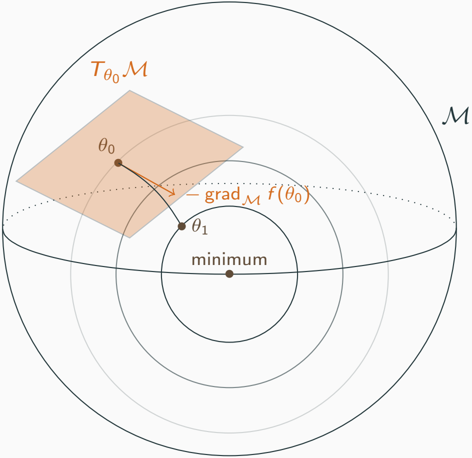
Optimization on a manifold



Optimization on a manifold



Optimization on a manifold



Step 2: objectives for feature estimation

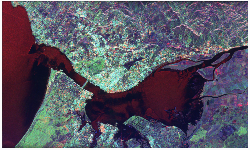


Figure 8: Example of a SAR image (from `nasa.gov`).

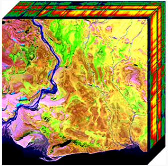


Figure 9: Example of a hyperspectral image (from `nasa.gov`).

Objectives:

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

Step 3: objectives for clustering/classification

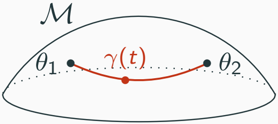


Figure 10: Distance: length of the geodesic γ.

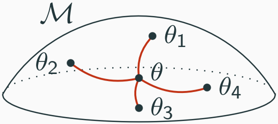


Figure 11: Center of mass: $\theta = \underset{\theta \in \mathcal{M}}{\arg\min} \sum_i d_\gamma^2(\theta, \theta_i)$.

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” statistical model

Study of a "low rank" statistical model

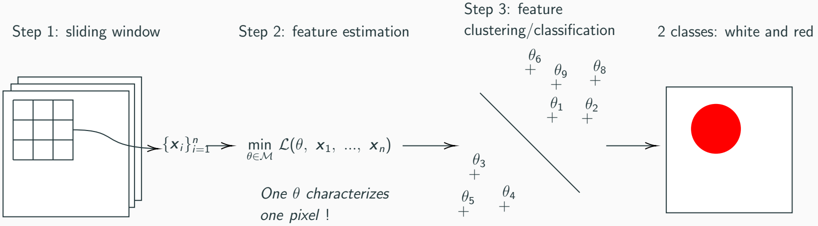


Figure 12: Clustering/classification pipeline.

Statistical model

$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{R}^p$, $\forall k < p$:

$$\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{0}, \tau_i \boldsymbol{U}\boldsymbol{U}^T + \boldsymbol{I}_p) \tag{3}$$

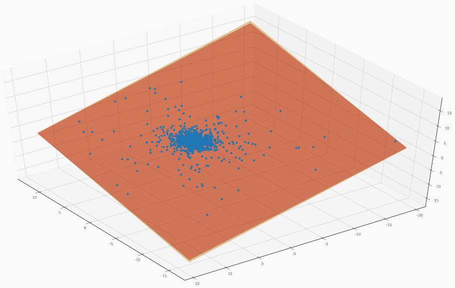
with $\tau_i > 0$ and $\boldsymbol{U} \in \mathbb{R}^{p \times k}$ is an orthogonal basis ($\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_k$).
Goal: estimate and classify $\theta = (\boldsymbol{U}, \boldsymbol{\tau})$.

Study of a "low rank" statistical model

Statistical model

$$\underbrace{\boldsymbol{x}_i}_{\in\mathbb{R}^p} \overset{d}{=} \underbrace{\sqrt{\tau_i}\boldsymbol{U}\boldsymbol{g}_i}_{\text{signal}\in\text{span}(\boldsymbol{U})} + \underbrace{\boldsymbol{n}_i}_{\text{noise}\in\mathbb{R}^p} \tag{4}$$

where $\boldsymbol{g}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_k)$ and $\boldsymbol{n}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_p)$ are independent, $\boldsymbol{\tau} \in (\mathbb{R}_*^+)^n$, and $\boldsymbol{U} \in \mathbb{R}^{p\times k}$ is an orthogonal basis ($\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_k$).



Study of a "low rank" statistical model: estimation

Maximum likelihood estimation (MLE)

Minimization of the negative log-likelihood with constraints:

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

$$\underset{(\boldsymbol{U},\boldsymbol{\tau})\in\mathrm{Gr}_{p,k}\times(\mathbb{R}_*^+)^n}{\text{minimize}} \ \mathcal{L}(\boldsymbol{U},\boldsymbol{\tau}) \tag{5}$$

Study of a "low rank" statistical model: estimation

Fisher information metric

$\forall \xi = (\boldsymbol{\xi_U}, \boldsymbol{\xi_\tau}), \eta = (\boldsymbol{\eta_U}, \boldsymbol{\eta_\tau})$ in the tangent space

$$\langle \xi, \eta \rangle^{\mathrm{FIM}}_{(\boldsymbol{U},\boldsymbol{\tau})} = \mathbb{E}\left[\mathrm{D}\,\mathcal{L}(\theta)[\xi]\,\mathrm{D}\,\mathcal{L}(\theta)[\eta]\right] \tag{6}$$

$$= 2nc_{\boldsymbol{\tau}}\,\mathrm{Tr}\left(\boldsymbol{\xi_U}^T \boldsymbol{\eta_U}\right) + k\left(\boldsymbol{\xi_\tau} \odot (\mathbf{1} + \boldsymbol{\tau})^{\odot -1}\right)^T \left(\boldsymbol{\eta_\tau} \odot (\mathbf{1} + \boldsymbol{\tau})^{\odot -1}\right), \tag{7}$$

where $c_{\boldsymbol{\tau}} = \frac{1}{n}\sum_{i=1}^n \frac{\tau_i^2}{1+\tau_i}$.

To solve (5) : Riemannian gradient descent on $\left(\mathrm{Gr}_{p,k} \times (\mathbb{R}_*^+)^n, \langle ., . \rangle^{\mathrm{FIM}}_{.}\right)$.

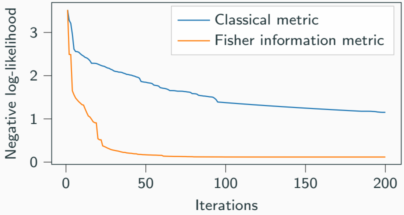


Figure 13: Negative log-likelihood versus the iterations.

Study of a "low rank" statistical model: bounds

Intrinsic Cramér-Rao bounds

Study of the performance through intrinsic Cramér-Rao bounds:

$$\overbrace{\mathbb{E}[d^2_{\mathrm{Gr}_{p,k}}(\mathrm{span}(\hat{\boldsymbol{U}}), \mathrm{span}(\boldsymbol{U}))]}^{\text{subspace estimation error}} \geq \frac{(p-k)k}{nc_{\boldsymbol{\tau}}} \approx \frac{(p-k)k}{n \times \mathrm{SNR}} \tag{8}$$

$$\underbrace{\mathbb{E}[d^2_{(\mathbb{R}_*^+)^n}(\hat{\boldsymbol{\tau}}, \boldsymbol{\tau})]}_{\text{texture estimation error}} \geq \frac{1}{k} \sum_{i=1}^{n} \frac{(1+\tau_i)^2}{\tau_i^2} \tag{9}$$

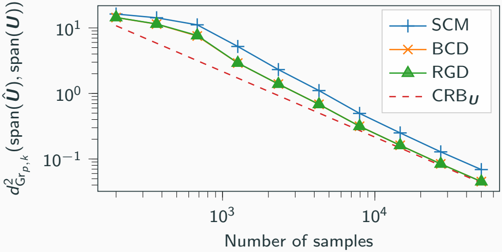


Figure 14: Mean squared error versus the number of simulated data.

Study of a "low rank" statistical model: K-means++

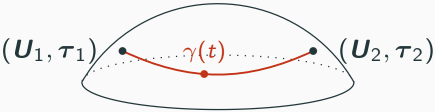


Figure 15: Distance.

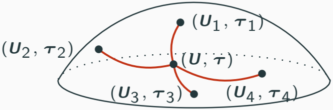


Figure 16: Center of mass (U, τ).

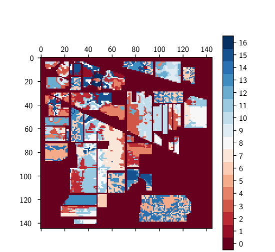


Figure 17: Euclidean *K-means++*: OA = 31.2%.

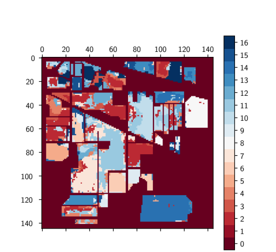


Figure 18: Proposed *K-means++*: OA = 47.2%.

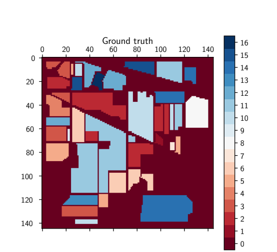


Figure 19: Ground truth.

Geodesic triangles for classification problems

Geodesic triangles for machine learning

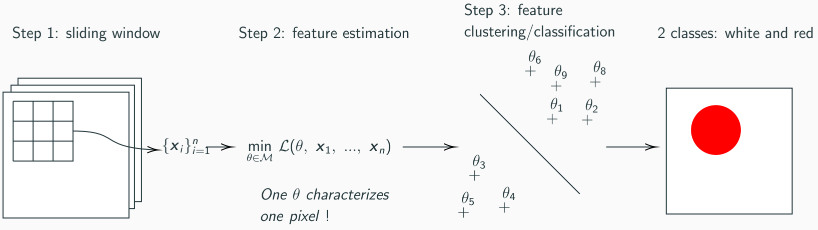


Figure 20: Clustering/classification pipeline.

Statistical model

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

Goal: classify $\theta = (\boldsymbol{\mu}, \Sigma)$.

Geodesic triangles for machine learning

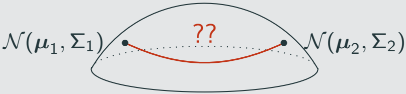
Riemannian geometry of Gaussian distributions

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

$$\langle \xi, \eta \rangle_{(\boldsymbol{\mu},\Sigma)}^{\mathrm{FIM}} = \boldsymbol{\xi}_{\boldsymbol{\mu}}^T \Sigma^{-1} \boldsymbol{\eta}_{\boldsymbol{\mu}} + \frac{1}{2} \operatorname{Tr}\left(\Sigma^{-1} \boldsymbol{\xi}_{\Sigma} \Sigma^{-1} \boldsymbol{\eta}_{\Sigma}\right). \tag{10}$$

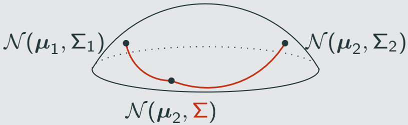
Problem

This Riemannian geometry is not fully known...



Geodesic triangles for machine learning

Solution: use of geodesic triangles



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

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

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

Riemannian center of mass $(\boldsymbol{\mu}, \Sigma)$ of $\{(\boldsymbol{\mu}_i, \Sigma_i)\}$

$$(\boldsymbol{\mu}, \Sigma) = \underset{(\boldsymbol{\mu}, \Sigma) \in \mathbb{R}^p \times \mathcal{S}_p^{++}}{\arg\min} \sum_i \delta^2\left((\boldsymbol{\mu}, \Sigma), (\boldsymbol{\mu}_i, \Sigma_i)\right) \tag{11}$$

A. Collas et al., "On The Use of Geodesic Triangles Between Gaussian Distributions for Classification Problems" in ICASSP 2022

Geodesic triangles for machine learning

Breizhcrops dataset:

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

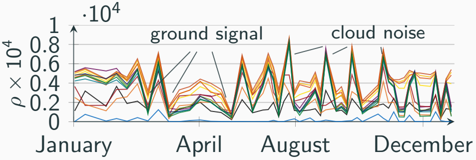


Figure 21: Reflectances of a Sentinel-2 time series.

Estimator	Geometry	Overall accuracy (%)
$\boldsymbol{X}$	$\mathbb{R}^{p \times n}$	10.1
sample mean $\hat{\boldsymbol{\mu}}$	$\mathbb{R}^p$	13.2
sample covariance matrix $\hat{\Sigma}$	$\mathcal{S}_p^{++}$	46.7
Proposed - $(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$	δ_c	**54.3**
Proposed - $(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$	$\delta_\perp$	**53.3**

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

Robust Geometric Metric Learning

Robust Geometric Metric Learning (RGML)

Let be a classification problem with K classes.

Metric learning

Find a *Mahalanobis* distance

$$d_{\boldsymbol{A}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{A}^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j)} \tag{12}$$

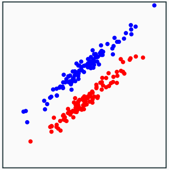
relevant for classification problems.

Metric learning as covariance estimation

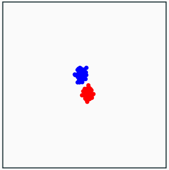
Proposed minimization problem:

$$\underset{(\boldsymbol{A}, \{\boldsymbol{A}_k\}) \in \left(\mathcal{S}_p^{++}\right)^{K+1}}{\text{minimize}} \quad \underbrace{\sum_{k=1}^{K} \pi_k \mathcal{L}_k(\boldsymbol{A}_k)}_{\text{negative log-likelihood}} + \lambda \underbrace{\sum_{k=1}^{K} \pi_k d^2_{\mathcal{S}_p^{++}}(\boldsymbol{A}, \boldsymbol{A}_k)}_{\substack{\text{cost function to compute} \\ \text{the center of mass of } \{\boldsymbol{A}_k\}}} \tag{13}$$

$\{\pi_k\}$ are the proportions of the classes and $\{\mathcal{L}_k\}$ are to be defined.



$\{\boldsymbol{x}_i\}$



$\left\{\boldsymbol{A}^{-\frac{1}{2}} \boldsymbol{x}_i\right\}$

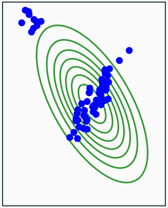
Robust Geometric Metric Learning (RGML)

Let $\boldsymbol{s}_{ki} = \boldsymbol{x}_l - \boldsymbol{x}_m$ where $\boldsymbol{x}_l, \boldsymbol{x}_m$ belong to the class k.

Gaussian negative log-likelihood

$$\mathcal{L}_{G,k}(\boldsymbol{A}_k) = \frac{1}{n_k} \sum_{i=1}^{n_k} \boldsymbol{s}_{ki}^T \boldsymbol{A}_k^{-1} \boldsymbol{s}_{ki} + \log |\boldsymbol{A}_k| \tag{14}$$

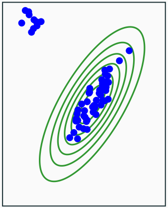
$$\text{minimized for } \boldsymbol{A}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \boldsymbol{s}_{ki} \boldsymbol{s}_{ki}^T \tag{15}$$



Tyler cost function

$$\mathcal{L}_{T,k}(\boldsymbol{A}_k) = \frac{p}{n_k} \sum_{i=1}^{n_k} \log \left(\boldsymbol{s}_{ki}^T \boldsymbol{A}_k^{-1} \boldsymbol{s}_{ki} \right) + \log |\boldsymbol{A}_k| \tag{16}$$

$$\text{minimized for } \boldsymbol{A}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \underbrace{\frac{p}{\boldsymbol{s}_{ki}^T \boldsymbol{A}_k^{-1} \boldsymbol{s}_{ki}}}_{\text{weight of sample } \boldsymbol{s}_{ki}} \boldsymbol{s}_{ki} \boldsymbol{s}_{ki}^T \tag{17}$$



Robust Geometric Metric Learning (RGML)

Riemannian metric

$\forall \xi = (\boldsymbol{\xi}, \{\boldsymbol{\xi}_k\}), \eta = (\boldsymbol{\eta}, \{\boldsymbol{\eta}_k\})$ in the tangent space

$$\langle \xi, \eta \rangle_{(\boldsymbol{A}, \{\boldsymbol{A}_k\})} = \mathrm{Tr}\left(\boldsymbol{A}^{-1}\boldsymbol{\xi}\boldsymbol{A}^{-1}\boldsymbol{\eta}\right) + \sum_{k=1}^{K} \mathrm{Tr}\left(\boldsymbol{A}_k^{-1}\boldsymbol{\xi}_k\boldsymbol{A}_k^{-1}\boldsymbol{\eta}_k\right) \tag{18}$$

$\implies$ strongly geodesically convexity of the minimization problem
$\implies$ the Riemannian gradient descent is fast

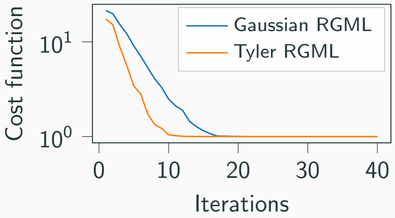


Figure 22: Cost function versus the iterations.

Robust Geometric Metric Learning (RGML)

RGML + *k-NN* on datasets from the UCI Machine Learning Repository

	Wine				Vehicle				Iris			
	$p = 13$, $n = 178$, $K = 3$				$p = 18$, $n = 846$, $K = 4$				$p = 4$, $n = 150$, $K = 3$			
Method	Mislabeling rate				Mislabeling rate				Mislabeling rate			
	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
Euclidean	30.12	30.40	31.40	32.40	38.27	38.58	39.46	40.35	3.93	4.47	5.31	**6.70**
SCM	10.03	11.62	13.70	17.57	23.59	24.27	25.24	26.51	12.57	13.38	14.93	16.68
ITML - Identity	3.12	4.15	5.40	**7.74**	24.21	23.91	24.77	26.03	3.04	4.47	5.31	**6.70**
ITML - SCM	2.45	4.76	6.71	10.25	23.86	23.82	24.89	26.30	3.05	13.38	14.92	16.67
GMML	2.16	3.58	5.71	9.86	21.43	22.49	23.58	25.11	2.60	5.61	9.30	12.62
LMNN	4.27	6.47	7.83	9.86	20.96	24.23	26.28	28.89	3.53	9.59	11.19	12.22
Proposed - Gaussian	**2.07**	**2.93**	5.15	9.20	**19.76**	21.19	22.52	24.21	**2.47**	5.10	8.90	12.73
Proposed - Tyler	**2.12**	**2.90**	**4.51**	8.31	19.90	**20.96**	**22.11**	**23.58**	**2.48**	**2.96**	**4.65**	7.83

Table 2: Misclassification errors on 3 datasets: Wine, Vehicle and Iris. Mislabeling rate: percentage of labels randomly changed in the training set.

Github: `https://github.com/antoinecollas/robust_metric_learning`

Conclusion

Conclusion

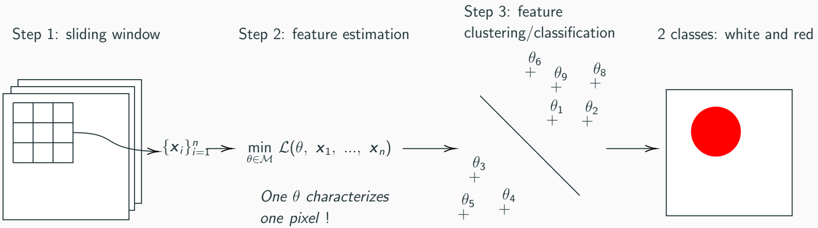


Figure 23: Clustering/classification pipeline.

Theoretical contributions:

- new statistical estimators, new Cramér-Rao bounds, new divergences, new Riemannian centers of mass, ...
- new strongly g-convex optimization problem in *metric learning*.

Applications on real datasets: *Indian pines* hyperspectral image, *Breizhcrops* multispectral times series, datasets from the *UCI* repository.

Conclusion

Software:

- contributor of *pyManopt*[2]: Python library for optimization on Riemannian manifolds,
- repository of *Robust Geometric Metric Learning*[3].

Near future:

- release *pyCovariance*: Python library for statistical estimation and clustering/classification using Riemannian geometry,

Future work:

- more sophisticated classifiers *e.g.* take into account the variance of θ (LDA/QDA on manifold)
- connection between optimal transport and optimization on manifold: continue the work done with Bamdev Mishra, creator of Manopt, in summer 2021
- apply these methods on other data such as EEG

[2] https://github.com/pymanopt/pymanopt
[3] https://github.com/antoinecollas/robust_metric_learning

Publications

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". In: *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". In: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

A. Collas, A. Breloy, G. Ginolhac, C. Ren, and J.-P. Ovarlez. "Robust Geometric Metric Learning". In: *Submitted to 2022 30th European Signal Processing Conference (EUSIPCO)*. 2022.

Journals:

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". In: *IEEE Transactions on Signal Processing* 69 (2021), pp. 6546–6560. DOI: 10.1109/TSP.2021.3130997.

Mian, A., **A. Collas**, A. Breloy, G. Ginolhac, and J.-P. Ovarlez. "Robust Low-Rank Change Detection for Multivariate SAR Image Time Series". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), pp. 3545–3556. DOI: 10.1109/JSTARS.2020.2999615.

Journal in preparation:

A. Collas, A. Breloy, C. Ren, G. Ginolhac, and J.-P. Ovarlez. "Riemannian Classification Approach To Non-Centered Mixture Of Scaled Gaussian".

Optimization and statistical learning using Riemannian geometry and application to remote sensing

Antoine Collas
PhD begun in October 2019, defense expected in September 2022
SONDRA lab, CentraleSupélec, Univ. Paris-Saclay

Supervisors:
Jean-Philippe Ovarlez - Director - ONERA/SONDRA
Guillaume Ginolhac - Director - LISTIC
Chengfang Ren - SONDRA
Arnaud Breloy - LEME

Riemannian geometry, robust estimation and libraries

Optimization on Riemannian manifolds:

Absil, P.-A., R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press, 2008, pp. xvi+224. ISBN: 978-0-691-13298-3.

Boumal, Nicolas. *An introduction to optimization on smooth manifolds*. Mar. 2022. URL: http://www.nicolasboumal.net/book.

Robust statistics:

Maronna, Ricardo Antonio. *Robust M-Estimators of Multivariate Location and Scatter*. 1976. DOI: 10.1214/aos/1176343347.

Ollila, Esa, David E. Tyler, Visa Koivunen, and H. Vincent Poor. *Complex Elliptically Symmetric Distributions: Survey, New Results and Applications*. 2012. DOI: 10.1109/TSP.2012.2212433.

Tyler, David E. *A Distribution-Free M-Estimator of Multivariate Scatter*. 1987. DOI: 10.1214/aos/1176350263.

Software:

Boumal, N., B. Mishra, P.-A. Absil, and R. Sepulchre. *Manopt, a Matlab Toolbox for Optimization on Manifolds*. 2014. URL: https://www.manopt.org.

Townsend, J., N. Koep, and S. Weichwald. *Pymanopt: A Python Toolbox for Optimization on Manifolds Using Automatic Differentiation*. Jan. 2016.

Study of a "low rank" statistical model: estimation

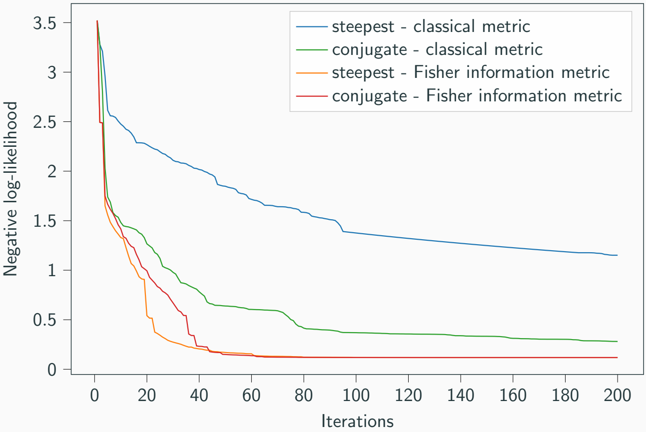


Figure 24: Negative log-likelihood versus the iterations.