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

WORKSHOP SONDRA - Avignon - 2022

Antoine Collas[1]

Work done with

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

Published in ICASSP 2022

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...).

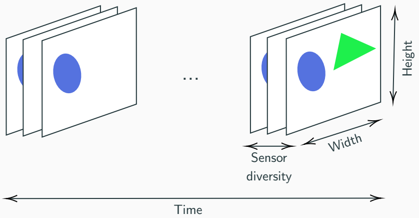


Figure 1: Multivariate image time series.

Applications

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

Example of multi-spectral time series

Breizhcrops dataset[1] [6]:

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

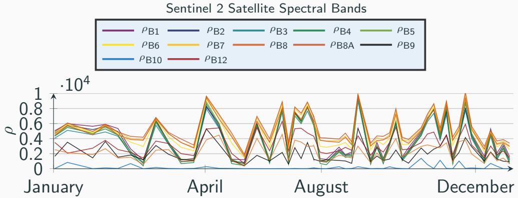


Figure 2: Reflectances ρ of a time series of **meadows**.

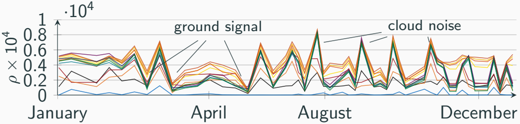


Figure 3: Reflectances ρ of a time series of **corn**.

[1] https://breizhcrops.org/

Clustering/classification pipeline and Riemannian geometry

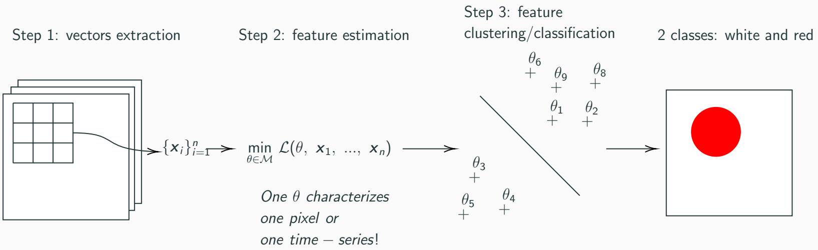


Figure 4: Clustering/classification pipeline.

Examples of θ:

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

Riemannian geometry and optimization

What is a Riemannian manifold ?

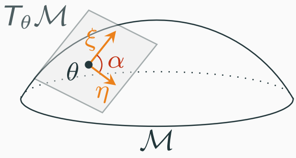


Figure 5: A Riemannian manifold.

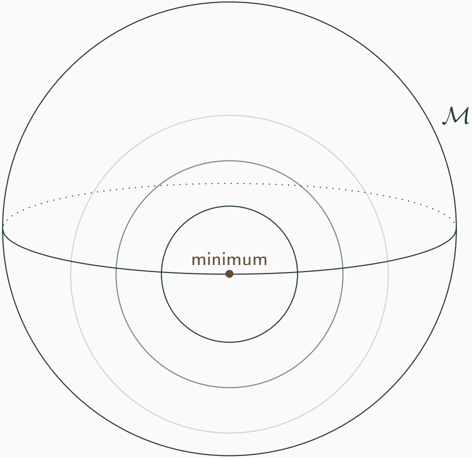
Curvature induced by:

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

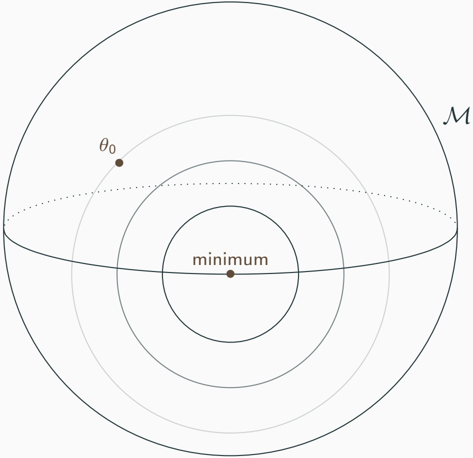
Examples of Riemannian manifolds $\mathcal{M}$:

- 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: $\mathcal{S}_p^{++} = \{\boldsymbol{\Sigma} \in \mathcal{S}_p : \forall \boldsymbol{x} \neq \boldsymbol{0} \in \mathbb{R}^p,\ \boldsymbol{x}^T\boldsymbol{\Sigma}\boldsymbol{x} > 0\}$
- norm constraints: $S^{p^2-1} = \{\boldsymbol{X} \in \mathbb{R}^{p\times p} : \|\boldsymbol{X}\|_F = 1\}$
- ...

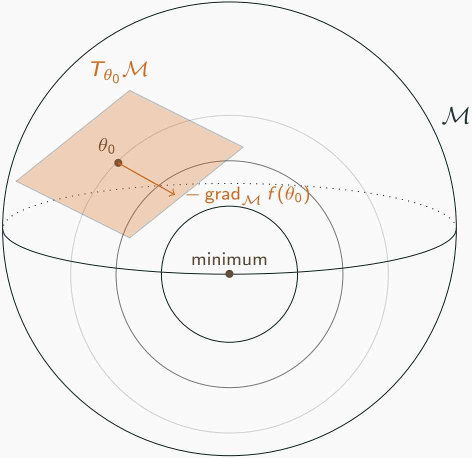
Optimization on a manifold



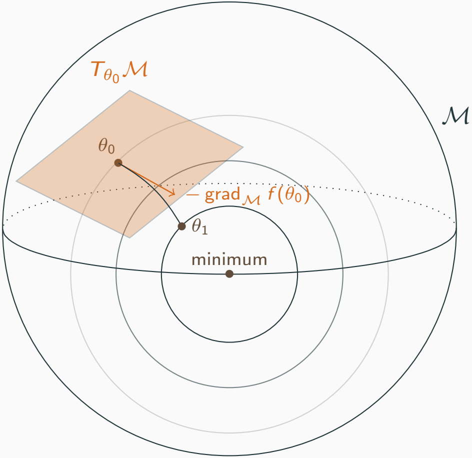
Optimization on a manifold



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

Step 2: maximum likelihood estimator

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

Step 3: Riemannian geometry of centered Gaussian distributions

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

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

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

Invariant by affine transformations:

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

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

Existing work (2/2)

Step 3

Riemannian distance between $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ in $\mathcal{S}_p^{++}$:

$$d_{\mathcal{S}_p^{++}}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \left\| \log \left(\mathbf{\Sigma}_1^{-\frac{1}{2}} \mathbf{\Sigma}_2 \mathbf{\Sigma}_1^{-\frac{1}{2}} \right) \right\|_2 . \tag{4}$$

Invariant by affine transformations:

$$d_{\mathcal{S}_p^{++}}(\phi(\mathbf{\Sigma}_1), \phi(\mathbf{\Sigma}_2)) = d_{\mathcal{S}_p^{++}}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \tag{5}$$

Riemannian center of mass of $\{\mathbf{\Sigma}_i\}_{i=1}^n$:

$$\mathbf{\Sigma} = \underset{\mathbf{\Sigma} \in \mathcal{S}_p^{++}}{\arg\min} \sum_{i=1}^{n} d^2_{\mathcal{S}_p^{++}}(\mathbf{\Sigma}, \mathbf{\Sigma}_i). \tag{6}$$

Enough to apply a *K-means* or a *Nearest centroïd classifier* on $\mathcal{S}_p^{++}$.

Geodesic triangles for classification problems

Geodesic triangles for machine learning

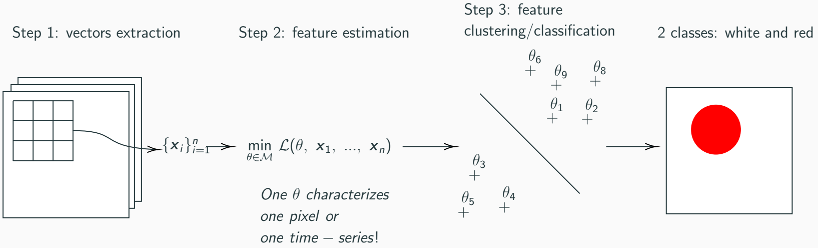


Figure 6: Clustering/classification pipeline.

Statistical model

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

Riemannian geometry of Gaussian distributions

Riemannian geometry of non-centered Gaussian distributions

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

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

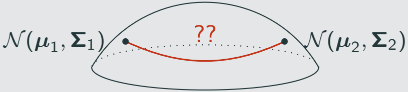
Invariant by affine transformations:

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

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

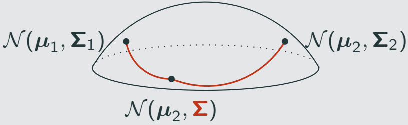
Problem

Problem: this Riemannian geometry is not fully known...



Geodesic triangles for machine learning

Solution: use of geodesic triangles



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

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

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

Invariant by affine transformations:

$$\delta\left(\phi(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \phi(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)\right) = \delta\left((\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)\right).$$

Riemannian center of mass

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

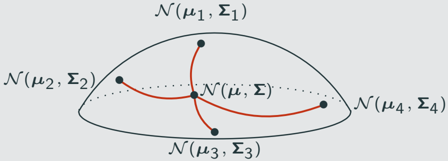


Figure 7: Center of mass $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \underset{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathbb{R}^p \times \mathcal{S}_p^{++}}{\arg\min} \left\{ f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_i \delta\left((\boldsymbol{\mu}, \boldsymbol{\Sigma}), (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\right) \right\}. \tag{9}$$

Riemannian gradient descent on $\mathbb{R}^p \times \mathcal{S}_p^{++}$

Iterations:

$$\left(\boldsymbol{\mu}_{k+1}, \boldsymbol{\Sigma}_{k+1}\right) := R_{(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}(-\alpha \,\mathrm{grad}\, f(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)) \tag{10}$$

where α is a step size, R a retraction and grad f the Riemannian gradient.

Machine learning: Nearest centroïd classifier on $\mathbb{R}^p \times \mathcal{S}_p^{++}$

Training

Labeled training data $\{(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), y_i\}_{i=1}^n$ with M classes.
Compute the Riemannian center of mass of each class: $\{(\boldsymbol{\mu}^m, \boldsymbol{\Sigma}^m)\}_{m=1}^M$.

Test

To classify the data $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: minimum divergence to mean:

$$m = \underset{m \in [\![1, M]\!]}{\arg\min} \left\{ \delta((\boldsymbol{\mu}^1, \boldsymbol{\Sigma}^1), (\boldsymbol{\mu}, \boldsymbol{\Sigma})), \cdots, \delta((\boldsymbol{\mu}^M, \boldsymbol{\Sigma}^M), (\boldsymbol{\mu}, \boldsymbol{\Sigma})) \right\} \quad (11)$$

Breizhcrops dataset [6]:

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

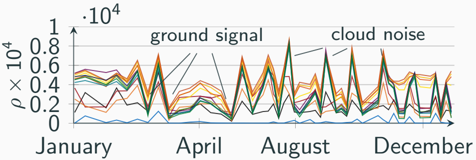


Figure 8: Reflectances of a Sentinel-2 time series.

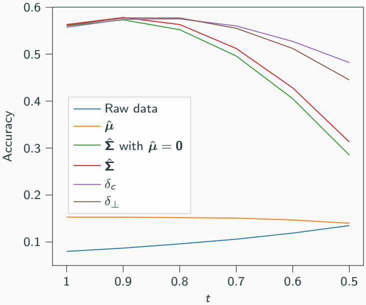
Estimation on each time series

- mean vector $\hat{\boldsymbol{\mu}} = \sum_{i=1}^{n} \boldsymbol{x}_i$,
- covariance matrix $\hat{\boldsymbol{\Sigma}} = \sum_{i=1}^{n} (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^T$.

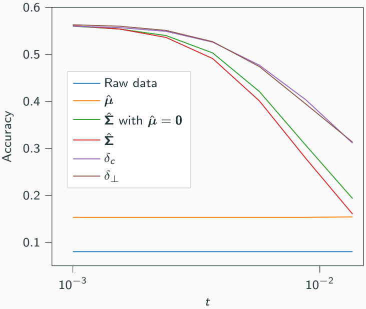
Machine learning

To show the robustness of the proposed method:

- training on raw data,
- test on transformed data.



(a) Scale transformation: $t \times \boldsymbol{x}_i$



(b) Rotation transformation: $\boldsymbol{Q}(t)\boldsymbol{x}_i$ with $\boldsymbol{Q}(t)$ a rotation matrix such that $\boldsymbol{Q}(0) = \boldsymbol{I}_p$

Conclusion

Conclusion

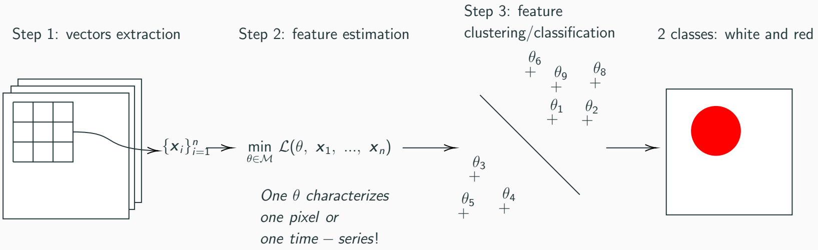


Figure 10: Clustering/classification pipeline.

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

WORKSHOP SONDRA - Avignon - 2022

Antoine Collas[1]

Work done with

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

Published in ICASSP 2022

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

[2] 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`.

[3] X. Pennec, P. Fillard, and N. Ayache, “A Riemannian framework for tensor computing,” *International Journal of computer vision*, vol. 66, no. 1, pp. 41–66, 2006.

[4] 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.

[5] 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.

[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.

Riemannian optimization

Proposition (Riemannian gradient)

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

where $\operatorname{grad}_{\epsilon} f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\boldsymbol{G}_{\boldsymbol{\mu}}, \boldsymbol{G}_{\boldsymbol{\Sigma}})$ *is the Euclidean gradient.*

Proposition (Second order retraction)

Given $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and* $\left(\xi_{\boldsymbol{\mu}}, \xi_{\boldsymbol{\Sigma}}\right)$ *in the tangent space*

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