# Apprentissage robuste de distance par géométrie riemannienne

GRETSI 2022 - Nancy, France

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

1 - SONDRA, CentraleSupélec, Université Paris Saclay
2 - LEME, Université Paris Nanterre
3 - LISTIC, Université Savoie Mont-Blanc
4 - DEMR, ONERA, Université Paris Saclay

# Table of contents

# Metric learning

# Metric learning

Supervised regime with $K$ classes: $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$.

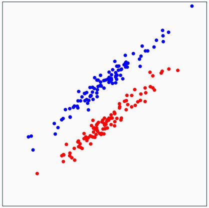
**Metric learning**

Find a *Mahalanobis* distance

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

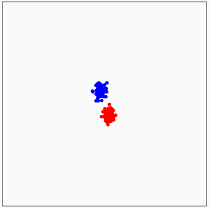
relevant for classification problems.

$\boldsymbol{A} \in \mathcal{S}_p^{++}$ the set of $p \times p$ symmetric positive definite matrices.



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

**(a)** Raw data



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

**(b)** Whitened data

# Information-Theoretic Metric Learning (ITML)

Set $S$: $n_S$ pairs $(\boldsymbol{x}_l, \boldsymbol{x}_q)$ with $y_l = y_q$.
Set $D$: $n_D$ pairs $(\boldsymbol{x}_l, \boldsymbol{x}_q)$ with $y_l \neq y_q$.

**ITML minimization problem**

Given $\boldsymbol{A}_0 \in \mathcal{S}_p^{++}$, and $u, v > 0$

$$\begin{aligned} \underset{\boldsymbol{A} \in \mathcal{S}_p^{++}}{\text{minimize}} \quad & \operatorname{Tr}(\boldsymbol{A}^{-1}\boldsymbol{A}_0) + \log|\boldsymbol{A}| \\ \text{subject to} \quad & d_{\boldsymbol{A}}^2(\boldsymbol{x}_l, \boldsymbol{x}_q) \leq u, \quad (\boldsymbol{x}_l, \boldsymbol{x}_q) \in S \\ & d_{\boldsymbol{A}}^2(\boldsymbol{x}_l, \boldsymbol{x}_q) \geq v, \quad (\boldsymbol{x}_l, \boldsymbol{x}_q) \in D \end{aligned}$$

**Interpreted as a covariance estimation problem**

For $\boldsymbol{A}_0 = \frac{1}{m}\sum_{i=1}^m \boldsymbol{x}_i \boldsymbol{x}_i^T$, it is the minimization of the Gaussian negative log-likelihood under constraints.

# Geometric Mean Metric Learning (GMML) (1/2)

## Minimization problem

$$\underset{\boldsymbol{A} \in \mathcal{S}_p^{++}}{\text{minimize}} \frac{1}{n_S} \sum_{(\boldsymbol{x}_l, \boldsymbol{x}_q) \in S} d^2_{\boldsymbol{A}}(\boldsymbol{x}_l, \boldsymbol{x}_q) + \frac{1}{n_D} \sum_{(\boldsymbol{x}_l, \boldsymbol{x}_q) \in D} d^2_{\boldsymbol{A}^{-1}}(\boldsymbol{x}_l, \boldsymbol{x}_q)$$

## GMML Algorithm

$$\boldsymbol{A}^{-1} = \boldsymbol{S}^{-1} \#_t \boldsymbol{D} = \boldsymbol{S}^{-\frac{1}{2}} \left( \boldsymbol{S}^{\frac{1}{2}} \boldsymbol{D} \boldsymbol{S}^{\frac{1}{2}} \right)^t \boldsymbol{S}^{-\frac{1}{2}} \text{ with } t \in [0, 1]$$

$\boldsymbol{S} = \frac{1}{n_S} \sum_{(\boldsymbol{x}_l, \boldsymbol{x}_q) \in S} (\boldsymbol{x}_l - \boldsymbol{x}_q)(\boldsymbol{x}_l - \boldsymbol{x}_q)^T$
$\boldsymbol{D} = \frac{1}{n_D} \sum_{(\boldsymbol{x}_l, \boldsymbol{x}_q) \in D} (\boldsymbol{x}_l - \boldsymbol{x}_q)(\boldsymbol{x}_l - \boldsymbol{x}_q)^T$.

$\boldsymbol{A}^{-1}$ is the Riemannian interpolation on $\mathcal{S}_p^{++}$ between $\boldsymbol{S}^{-1}$ and $\boldsymbol{D}$.

In practice, works well for $t$ small, *i.e.* $\boldsymbol{A} \approx \boldsymbol{S}$.

# Geometric Mean Metric Learning (GMML) (2/2)

**Assumption**

*Data points of each class are realizations of independent random vectors with class-dependent first and second order moments*

$$\boldsymbol{x}_{kl} \stackrel{d}{=} \boldsymbol{\mu}_k + \boldsymbol{\Sigma}_k^{\frac{1}{2}} \boldsymbol{u}_{kl} \, ,$$

*with* $\boldsymbol{\mu}_k \in \mathbb{R}^p$, $\boldsymbol{\Sigma}_k \in \mathcal{S}_p^{++}$, $\mathbb{E}[\boldsymbol{u}_{kl}] = \boldsymbol{0}$ *and* $\mathbb{E}[\boldsymbol{u}_{kl}\boldsymbol{u}_{kq}^T] = \boldsymbol{I}$ *if* $kl = kq$, $\boldsymbol{0}_p$ *otherwise.*

**Interpreted as a covariance estimation problem**

$$\mathbb{E}[\boldsymbol{S}] = 2 \sum_{k=1}^{K} \pi_k \boldsymbol{\Sigma}_k$$

where $\{\pi_k\}$ are the classes proportions.

Thus, in practice

$$\mathbb{E}[\boldsymbol{A}] \approx 2 \sum_{k=1}^{K} \pi_k \boldsymbol{\Sigma}_k .$$

# Robust Geometric Metric Learning

# Robust Geometric Metric Learning (RGML)

**Proposed general formulation**

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

where $d_{\mathcal{S}_p^{++}}$ is the Riemannian distance on $\mathcal{S}_p^{++}$

$$d^2_{\mathcal{S}_p^{++}}\left(\boldsymbol{A},\boldsymbol{A}_k\right) = \left\|\operatorname{logm}\left(\boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{A}_k\boldsymbol{A}^{-\frac{1}{2}}\right)\right\|_F^2 .$$

# Robust Geometric Metric Learning (RGML)

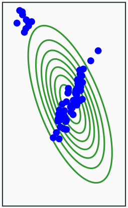
Set $S_k$: $n_k$ pairs $(\boldsymbol{x}_l, \boldsymbol{x}_q)$ with $y_l = y_q = k$.

## Gaussian negative log-likelihood

$$\mathcal{L}_{G,k}(\boldsymbol{A}_k) = \frac{1}{n_k} \sum_{(\boldsymbol{x}_l, \boldsymbol{x}_q) \in S_k} (\boldsymbol{x}_l - \boldsymbol{x}_q)^T \boldsymbol{A}_k^{-1} (\boldsymbol{x}_l - \boldsymbol{x}_q) + \log |\boldsymbol{A}_k|$$

minimized for

$$\boldsymbol{A}_k = \frac{1}{n_k} \sum_{(\boldsymbol{x}_l, \boldsymbol{x}_q) \in S_k} (\boldsymbol{x}_l - \boldsymbol{x}_q)(\boldsymbol{x}_l - \boldsymbol{x}_q)^T$$

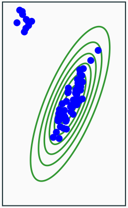


## Tyler cost function

$$\mathcal{L}_{T,k}(\boldsymbol{A}_k) = \frac{p}{n_k} \sum_{(\boldsymbol{x}_l, \boldsymbol{x}_q) \in S_k} \log \left( (\boldsymbol{x}_l - \boldsymbol{x}_q)^T \boldsymbol{A}_k^{-1} (\boldsymbol{x}_l - \boldsymbol{x}_q) \right) + \log |\boldsymbol{A}_k|$$

minimized for

$$\boldsymbol{A}_k = \frac{1}{n_k} \sum_{(\boldsymbol{x}_l, \boldsymbol{x}_q) \in S_k} \underbrace{\frac{p}{(\boldsymbol{x}_l - \boldsymbol{x}_q)^T \boldsymbol{A}_k^{-1} (\boldsymbol{x}_l - \boldsymbol{x}_q)}}_{\text{weight of } (\boldsymbol{x}_l - \boldsymbol{x}_q)} (\boldsymbol{x}_l - \boldsymbol{x}_q)(\boldsymbol{x}_l - \boldsymbol{x}_q)^T$$



# Robust Geometric Metric Learning (RGML)

## Gaussian RGML

$$\underset{(\boldsymbol{A},\{\boldsymbol{A}_k\})\in\left(\mathcal{S}_p^{++}\right)^{K+1}}{\text{minimize}} h_G\left(\boldsymbol{A},\{\boldsymbol{A}_k\}\right) = \underbrace{\sum_{k=1}^{K} \pi_k \mathcal{L}_{G,k}(\boldsymbol{A}_k)}_{\text{Gaussian negative log-likelihood}} + \lambda \sum_{k=1}^{K} \pi_k d^2_{\mathcal{S}_p^{++}}(\boldsymbol{A}, \boldsymbol{A}_k)$$

## Tyler RGML

$$\underset{(\boldsymbol{A},\{\boldsymbol{A}_k\})\in\left(\mathcal{SS}_p^{++}\right)^{K+1}}{\text{minimize}} h_T\left(\boldsymbol{A},\{\boldsymbol{A}_k\}\right) = \underbrace{\sum_{k=1}^{K} \pi_k \mathcal{L}_{T,k}(\boldsymbol{A}_k)}_{\text{Tyler cost function}} + \lambda \sum_{k=1}^{K} \pi_k d^2_{\mathcal{S}_p^{++}}(\boldsymbol{A}, \boldsymbol{A}_k)$$

where $\mathcal{SS}_p^{++} = \left\{\boldsymbol{\Sigma} \in \mathcal{S}_p^{++} : |\boldsymbol{\Sigma}| = 1\right\}$

# Riemannian geometry and optimization

# What is a Riemannian manifold ?

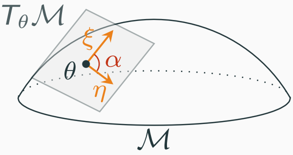



**Figure 2:** A Riemannian manifold.

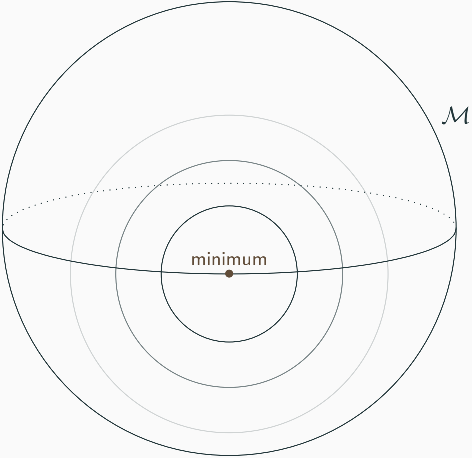
Curvature induced by:

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

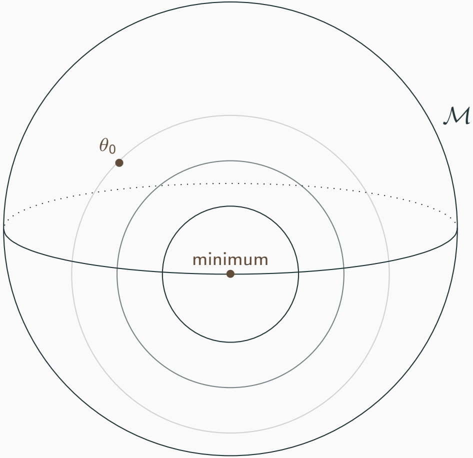
**Examples of Riemannian manifolds $\mathcal{M}$:**

- 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: $\mathcal{S}_p^{++} = \{\boldsymbol{\Sigma} \in \mathcal{S}_p : \forall \boldsymbol{x} \neq \boldsymbol{0} \in \mathbb{R}^p,\ \boldsymbol{x}^T\boldsymbol{\Sigma}\boldsymbol{x} > 0\}$
- positivity constraints: $\mathcal{SS}_p^{++} = \{\boldsymbol{\Sigma} \in \mathcal{S}_p^{++} : |\boldsymbol{\Sigma}| = 1\}$
- rank constraints: $\mathcal{S}_{p,k}^{+} = \{\boldsymbol{\Sigma} \in \mathcal{S}_p^{+} : \mathrm{rank}(\boldsymbol{\Sigma}) = k\}$
- norm constraints: $S^{p^2-1} = \{\boldsymbol{X} \in \mathbb{R}^{p \times p} : \|\boldsymbol{X}\|_F = 1\}$

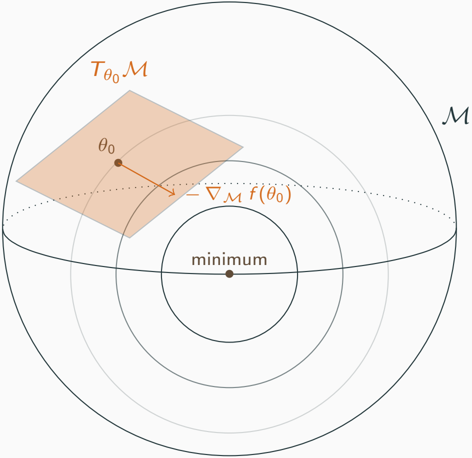
# Optimization on a manifold



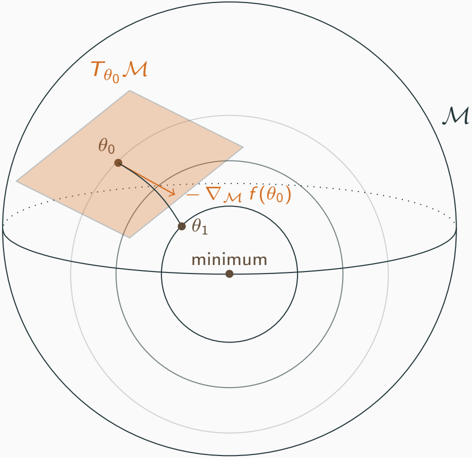
# Optimization on a manifold



# Optimization on a manifold



# Optimization on a manifold



# Robust Geometric Metric Learning (RGML)

**Chosen Riemannian metric on $\left(\mathcal{S}_p^{++}\right)^{K+1}$ and $\left(\mathcal{S}\mathcal{S}_p^{++}\right)^{K+1}$**

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

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

$\implies$ the minimization problems are strongly geodesically convex
$\implies$ unique global minimum and fast Riemannian gradient descent

# Robust Geometric Metric Learning (RGML)

$\theta = (\boldsymbol{A}, \{\boldsymbol{A}_k\})$, $\alpha$ a step size

**Iterations of Gaussian RGML: minimization of $h_G$**

$$\theta_{\ell+1} = \underbrace{R_{\theta_\ell}^{\left(\mathcal{S}_p^{++}\right)^{K+1}}}_{\text{retraction on } \left(\mathcal{S}_p^{++}\right)^{K+1}} \left( -\alpha \underbrace{\nabla^{\left(\mathcal{S}_p^{++}\right)^{K+1}} h_G(\theta_\ell)}_{\text{Riemannian gradient of } h_G} \right)$$

**Iterations of Tyler RGML: minimization of $h_T$**

$$\theta_{\ell+1} = \underbrace{R_{\theta_\ell}^{\left(\mathcal{S}\mathcal{S}_p^{++}\right)^{K+1}}}_{\text{retraction on } \left(\mathcal{S}\mathcal{S}_p^{++}\right)^{K+1}} \left( -\alpha \underbrace{\nabla^{\left(\mathcal{S}\mathcal{S}_p^{++}\right)^{K+1}} h_T(\theta_\ell)}_{\text{Riemannian gradient of } h_T} \right)$$

# Application

# Robust Geometric Metric Learning (RGML)

Application to datasets from the UCI Machine Learning Repository

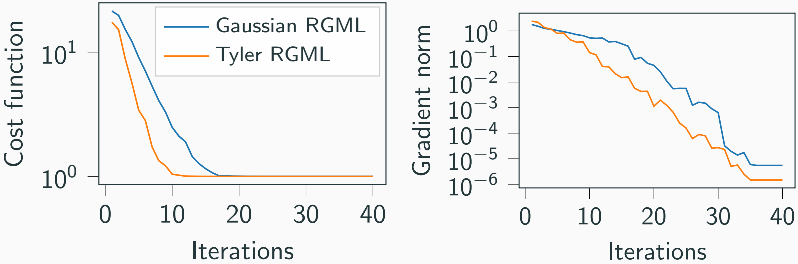



**Figure 3:** Left: cost function versus the number of iterations. Right: gradient norm versus the number of iterations. The optimization is performed on the *Wine* dataset.

# 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 1:** 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

**Theoretical contributions**:

- new interpretation of GMML algorithm...
- new g-convex optimization problem in *metric learning*: Gaussian RGML and Tyler RGML.

**Application** to real datasets from the *UCI* repository: RGML is fast, performant and robust to mislabelling.

# Publications


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

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

Davis, J. V., B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. "Information-Theoretic Metric Learning". In: *Proceedings of the 24th International Conference on Machine Learning*. ICML '07. Corvalis, Oregon, USA: Association for Computing Machinery, 2007, pp. 209–216. ISBN: 9781595937933. DOI: `10.1145/1273496.1273523`. URL: `https://doi.org/10.1145/1273496.1273523`.

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`.

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

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

# Apprentissage robuste de distance par géométrie riemannienne

GRETSI 2022 - Nancy, France

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

1 - SONDRA, CentraleSupélec, Université Paris Saclay
2 - LEME, Université Paris Nanterre
3 - LISTIC, Université Savoie Mont-Blanc
4 - DEMR, ONERA, Université Paris Saclay