

# Robust Geometric Metric Learning

---

EUSIPCO 2022 - Belgrade, Serbia

A. Collas<sup>1</sup>, A. Breloy<sup>2</sup>, G. Ginolhac<sup>3</sup>, C. Ren<sup>1</sup>, J.-P. Ovarlez<sup>1,4</sup>

1 - SONDRRA, CentraleSupélec, University Paris Saclay

2 - LEME, University Paris Nanterre

3 - LISTIC, University Savoie Mont-Blanc

4 - DEMR, ONERA, University Paris Saclay

# Table of contents

1. Metric learning
2. Robust Geometric Metric Learning
3. Riemannian geometry and optimization
4. Application

# Metric learning

---

# Metric learning

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

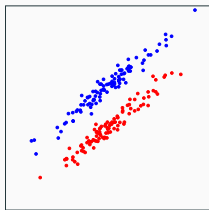
## Metric learning

Find a *Mahalanobis* distance

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

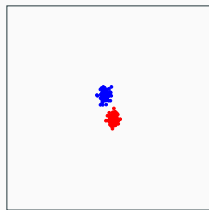
relevant for classification problems.

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



$\{\mathbf{x}_i\}$

(a) Raw data



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

(b) Whitened data

# Information-Theoretic Metric Learning (ITML)

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

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

## ITML minimization problem

Given  $\mathbf{A}_0 \in \mathcal{S}_p^{++}$

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

## Interpreted as a covariance estimation problem

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

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

## Minimization problem

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

## GMML Algorithm

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

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

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

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

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

# 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*

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

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

## Interpreted as a covariance estimation problem

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

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

Thus, in practice

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

# Robust Geometric Metric Learning

---



# Robust Geometric Metric Learning (RGML)

## Proposed general formulation

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

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

$$d_{\mathcal{S}_p^{++}}^2(\mathbf{A}, \mathbf{A}_k) = \left\| \log_m \left( \mathbf{A}^{-\frac{1}{2}} \mathbf{A}_k \mathbf{A}^{-\frac{1}{2}} \right) \right\|_F^2.$$

# Robust Geometric Metric Learning (RGML)

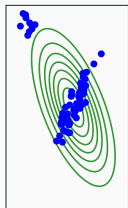
Set  $S_k$ :  $n_k$  pairs  $(\mathbf{x}_l, \mathbf{x}_q)$  with  $y_l = y_q = k$ .

## Gaussian negative log-likelihood

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

minimized for

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

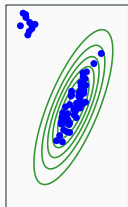


## Tyler cost function

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

minimized for

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



# Robust Geometric Metric Learning (RGML)

## Gaussian RGML

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

## Tyler RGML

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

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

# Riemannian geometry and optimization

---

# What is a Riemannian manifold ?

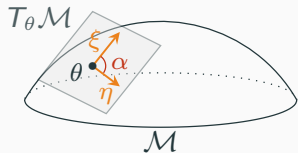


Figure 2: A Riemannian manifold.

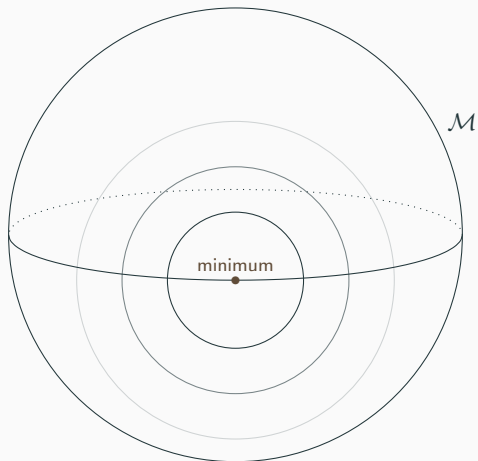
Curvature induced by:

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

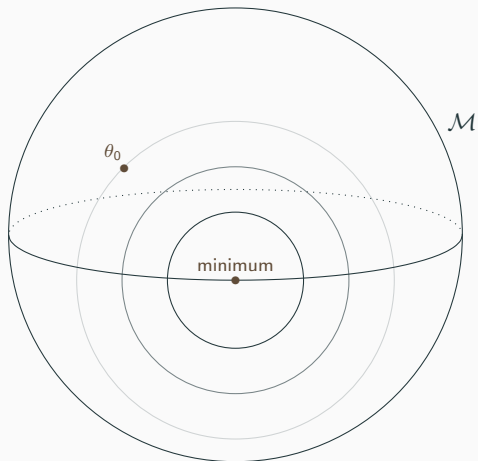
## Examples of Riemannian manifolds $\mathcal{M}$ :

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

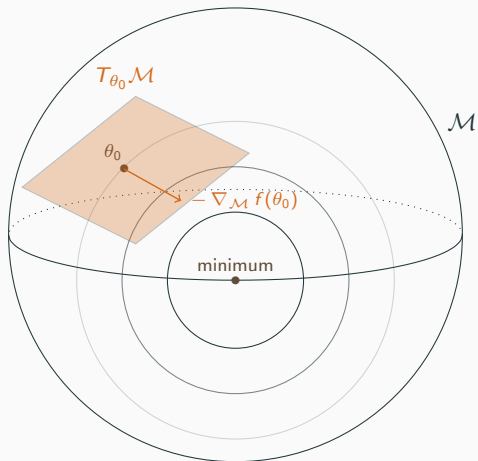
# Optimization on a manifold



# Optimization on a manifold

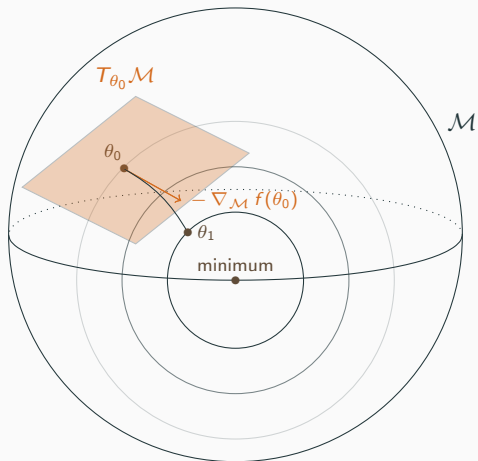


# Optimization on a manifold





# Optimization on a manifold



# Robust Geometric Metric Learning (RGML)

## Riemannian metric

$\forall \xi = (\mathbf{X}, \{\xi_k\}), \eta = (\boldsymbol{\eta}, \{\eta_k\})$  in the tangent space

$$\langle \xi, \eta \rangle_{(\mathbf{A}, \{\mathbf{A}_k\})} = \text{Tr}(\mathbf{A}^{-1} \xi \mathbf{A}^{-1} \eta) + \sum_{k=1}^K \text{Tr}(\mathbf{A}_k^{-1} \xi_k \mathbf{A}_k^{-1} \eta_k)$$

$\implies$  geodesically convexity of the minimization problems

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

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

$$\theta^{(k+1)} = \underbrace{R_{\theta^{(k)}}^{(S_p^{++})^{k+1}}}_{\text{retraction on } (S_p^{++})^{k+1}} \left( -\alpha \underbrace{\nabla h_G(\theta^{(k)})}_{\text{Riemannian gradient of } h_G} \right)$$

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

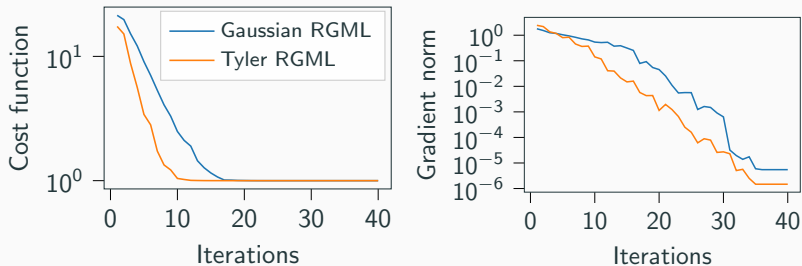
$$\theta^{(k+1)} = \underbrace{R_{\theta^{(k)}}^{(SS_p^{++})^{k+1}}}_{\text{retraction on } (SS_p^{++})^{k+1}} \left( -\alpha \underbrace{\nabla h_T(\theta^{(k)})}_{\text{Riemannian gradient of } h_T} \right)$$

# 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

Method	Wine $p = 13, n = 178, K = 3$				Vehicle $p = 18, n = 846, K = 4$				Iris $p = 4, n = 150, K = 3$			
	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	<b>6.70</b>
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	<b>7.74</b>	24.21	23.91	24.77	26.03	3.04	4.47	5.31	<b>6.70</b>
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	<b>2.07</b>	<b>2.93</b>	5.15	9.20	<b>19.76</b>	21.19	22.52	24.21	<b>2.47</b>	5.10	8.90	12.73
Proposed - Tyler	<b>2.12</b>	<b>2.90</b>	<b>4.51</b>	8.31	19.90	<b>20.96</b>	<b>22.11</b>	<b>23.58</b>	<b>2.48</b>	<b>2.96</b>	<b>4.65</b>	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](https://github.com/antoinecollas/robust_metric_learning)

## Conclusion

---

## Theoretical contributions:

- new interpretation of GMM 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.



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. Corvallis, 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.



# Robust Geometric Metric Learning

---

EUSIPCO 2022 - Belgrade, Serbia

A. Collas<sup>1</sup>, A. Breloy<sup>2</sup>, G. Ginolhac<sup>3</sup>, C. Ren<sup>1</sup>, J.-P. Ovarlez<sup>1,4</sup>

1 - SONDRRA, CentraleSupélec, University Paris Saclay

2 - LEME, University Paris Nanterre

3 - LISTIC, University Savoie Mont-Blanc

4 - DEMR, ONERA, University Paris Saclay