Probabilistic PCA from Heteroscedastic Signals: Geometric Framework and Application to Clustering

Antoine Collas
Work done with F. Bouchard, A. Breloy, C. Ren, G. Ginolhac, J.-P. Ovarlez

6 october 2021

Table of contents

Context

Context

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

Challenges

The objective is to develop clustering methods specific to these new data. More particularly we focus on 2 specific topics:

- Change detection.
- Semantic segmentation.

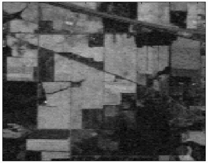


Figure 1: Raw image.

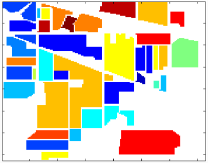


Figure 2: Segmented image. One color = one class (grass, woods, ...).

Objectives for parameter estimation

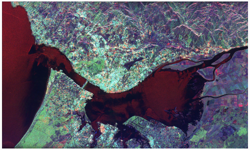


Figure 3: Example of a SAR image (from `nasa.gov`).

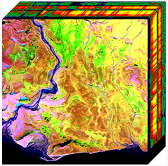


Figure 4: Example of a hyperspectral image (from `nasa.gov`).

Remark

To segment an image we begin with an estimation step. Because of the data, we have to develop:

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

Clustering pipeline and Riemannian geometry

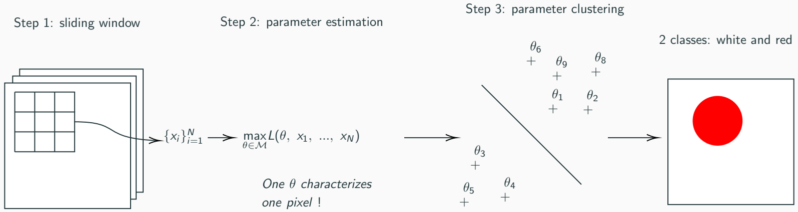


Figure 5: Clustering pipeline on an image.

The statistical model depends on $\theta \in \mathcal{M}$, a *structured parameter* in a *smooth manifold*.

- Step 2: maximization of the likelihood L over $\mathcal{M}$ which can be turned into a Riemannian geometry.
- Step 3: use of a Riemannian geometry of $\mathcal{M}$ to compute distances and means on $\mathcal{M}$.

Quick introduction on Riemannian geometry and optimization on matrix manifolds

Riemannian geometry

A Riemannian manifold is a couple $(\mathcal{M}, \langle \cdot, \cdot \rangle_\theta^\mathcal{M})$ where

- $\mathcal{M}$ is a *smooth manifold* (*i.e.* a locally Euclidean set),
- $\langle \cdot, \cdot \rangle_\theta^\mathcal{M}$ is an inner product, on $T_\theta \mathcal{M}$, called the *Riemannian metric*.

The vector space $T_\theta \mathcal{M}$ is called the tangent space and is the linearization of $\mathcal{M}$ at θ.

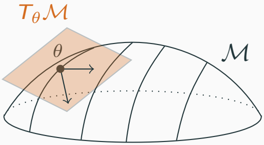


Figure 6: A manifold $\mathcal{M}$ with its tangent space $T_\theta \mathcal{M}$.

Introduction to optimization on matrix manifolds

Let f be a real-valued function to minimize over its parameter space:

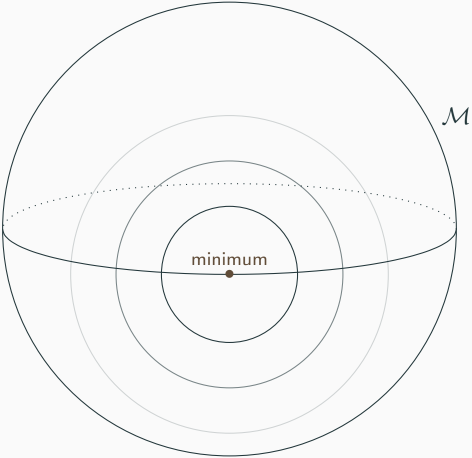
$$\min_{\theta \in \mathcal{M}} f(\theta) \tag{1}$$

where $\mathcal{M}$ is a Riemannian manifold which include the constraints of our problem.

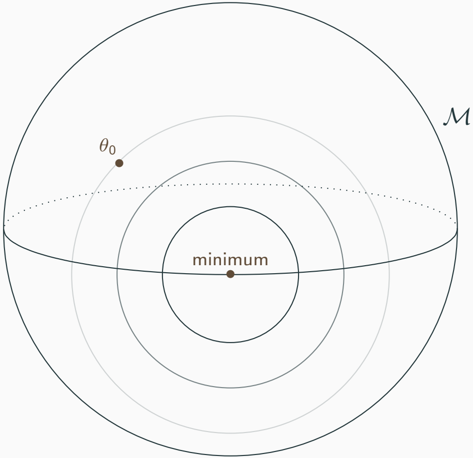
Examples of smooth manifolds $\mathcal{M}$:

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

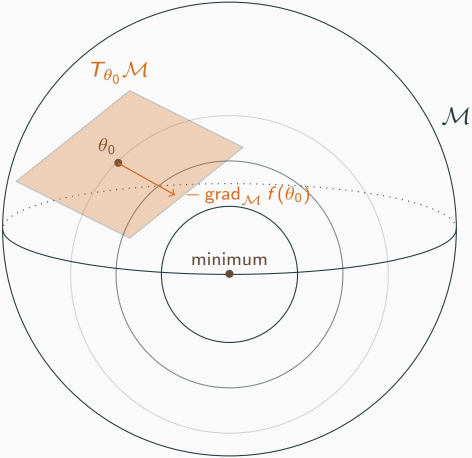
Introduction to optimization on matrix manifolds



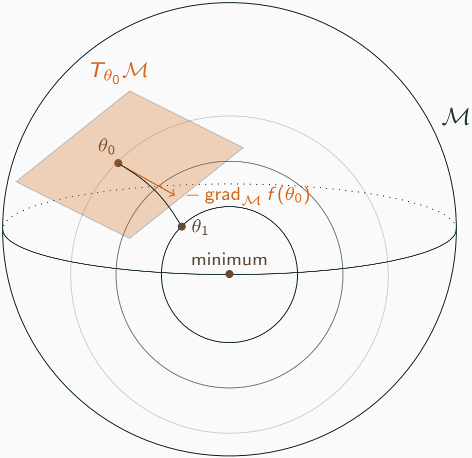
Introduction to optimization on matrix manifolds



Introduction to optimization on matrix manifolds



Introduction to optimization on matrix manifolds



Riemannian geometry and statistical estimation using the Fisher information metric

Data model (1/2)

$\forall k < p$, let n data points $\{\boldsymbol{x_i}\}_{i=1}^n \subset \mathbb{C}^p$ distributed as

$$\boldsymbol{x_i} \underset{d}{=} \sqrt{\tau_i}\boldsymbol{U}\boldsymbol{g_i} + \boldsymbol{n_i} \tag{2}$$

$\boldsymbol{g_i} \sim \mathbb{C}\mathcal{N}(0, \boldsymbol{I}_k)$, $\boldsymbol{n_i} \sim \mathbb{C}\mathcal{N}(0, \boldsymbol{I}_p)$ independent,
$\boldsymbol{\tau} \in (\mathbb{R}_*^+)^n$, $\boldsymbol{U} \in \mathrm{St}_{p,k} \triangleq \{\boldsymbol{U} \in \mathbb{C}^{p\times k} : \boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}_k\}$.

$$\implies \boldsymbol{x_i} \sim \mathbb{C}\mathcal{N}\left(0, \bar{\psi}_i(\bar{\theta}) \triangleq \boldsymbol{I}_p + \tau_i\boldsymbol{U}\boldsymbol{U}^H\right) \tag{3}$$

where $\bar{\theta} = (\boldsymbol{U}, \boldsymbol{\tau}) \in \overline{\mathcal{M}}_{p,k,n} \triangleq \mathrm{St}_{p,k} \times (\mathbb{R}_*^+)^n$.

Remark

For all $\boldsymbol{U} \in \mathrm{St}_{p,k}$ *and* $\boldsymbol{O} \in \mathcal{U}_k \triangleq \mathrm{St}_{k,k}$, $\bar{\psi}_i(\boldsymbol{U}\boldsymbol{O}, \boldsymbol{\tau}) = \bar{\psi}_i(\boldsymbol{U}, \boldsymbol{\tau})$.

Data model (2/2)

Definition the quotient parameter space

$$\mathcal{M}_{p,k,n} \triangleq \{\pi(\bar{\theta}) : \bar{\theta} \in \overline{\mathcal{M}}_{p,k,n}\} \text{ with } \pi(\bar{\theta}) = \{(\boldsymbol{UO}, \boldsymbol{\tau}) : \boldsymbol{O} \in \mathcal{U}_k\}. \quad (4)$$

Definition of the covariance matrix ψ_i on $\mathcal{M}_{p,k,n}$ from $\overline{\psi}_i$

$$\forall \theta = \pi(\bar{\theta}) \in \mathcal{M}_{p,k,n}, \quad \psi_i(\theta) \triangleq \overline{\psi}_i(\bar{\theta}) = \boldsymbol{I}_p + \tau_i \boldsymbol{U}\boldsymbol{U}^H. \quad (5)$$

The negative log-likelihood function is

$$L(\theta) \triangleq \bar{L}(\bar{\theta}) = \sum_{i=1}^{n} \left[\log\det \overline{\psi}_i(\theta) + \boldsymbol{x_i}^H (\overline{\psi}_i(\theta))^{-1} \boldsymbol{x_i}\right]. \quad (6)$$

Riemannian geometry of $\overline{\mathcal{M}}_{p,k,n}$

The tangent space of $\overline{\mathcal{M}}_{p,k,n}$ at $\bar{\theta} \in \overline{\mathcal{M}}_{p,k,n}$ is

$$T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k,n} = \{\bar{\xi} = (\boldsymbol{\xi_U}, \boldsymbol{\xi_\tau}) \in \mathbb{C}^{p\times k} \times \mathbb{R}^n : \ \boldsymbol{U}^H \boldsymbol{\xi_U} + \boldsymbol{\xi_U}^H \boldsymbol{U} = 0\}. \quad (7)$$

$\forall \bar{\xi}, \bar{\eta} \in T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k,n}$ the Fisher Information Metric is defined as

$$\langle \bar{\xi}, \bar{\eta} \rangle_{\bar{\theta}}^{\mathrm{FIM}} = \mathbb{E}[\mathrm{D}\,\bar{L}(\bar{\theta})[\bar{\xi}]\,\mathrm{D}\,\bar{L}(\bar{\theta})[\bar{\eta}]]. \quad (8)$$

Proposition (Fisher information metric)

The Fisher information metric on $\overline{\mathcal{M}}_{p,k,n}$ corresponding to the log-likelihood (6) *is*

$$\langle \bar{\xi}, \bar{\eta} \rangle_{\bar{\theta}}^{\mathrm{FIM}} = 2nc_{\boldsymbol{\tau}} \mathfrak{Re}(\mathrm{Tr}(\boldsymbol{\xi_U}^H \boldsymbol{\eta_U})) + k(\boldsymbol{\xi_\tau} \odot (1+\boldsymbol{\tau})^{\odot -1})^T (\boldsymbol{\eta_\tau} \odot (1+\boldsymbol{\tau})^{\odot -1}) \quad (9)$$

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

Parameter estimation: retraction and Riemannian gradient

$$\min_{\theta \in \mathcal{M}_{p,k,n}} \quad L(\theta) = \sum_{i=1}^{n} L_i(\theta) \tag{10}$$

We define a retraction: $T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k,n} \to \overline{\mathcal{M}}_{p,k,n}$:

$$\overline{R}_{\bar{\theta}}(\bar{\xi}) = \left(\boldsymbol{X}\boldsymbol{Y}^H, \boldsymbol{\tau} + \boldsymbol{\xi}_{\boldsymbol{\tau}} + \frac{1}{2}\boldsymbol{\tau}^{\odot -1}\boldsymbol{\xi}_{\boldsymbol{\tau}}^{\odot 2} \right) \tag{11}$$

where $\boldsymbol{U} + \boldsymbol{\xi}_{\boldsymbol{U}} = \boldsymbol{X}\Sigma\boldsymbol{Y}^H$ by SVD.

Definition of the Riemannian gradient:

$$\forall \bar{\xi} \in T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k,n}, \quad \mathrm{D}\,\bar{L}_i(\bar{\theta})[\bar{\xi}] = \langle \mathrm{grad}\,\bar{L}_i(\bar{\theta}), \bar{\xi} \rangle_{\bar{\theta}}^{\mathrm{FIM}}. \tag{12}$$

The representative in $T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k,n}$ of the Riemannian gradient of L_i at θ is

$$\mathrm{grad}\,\bar{L}_i(\bar{\theta}) = (\boldsymbol{G}_{\boldsymbol{U}}, \boldsymbol{G}_{\boldsymbol{\tau}}) \tag{13}$$

$$\boldsymbol{G}_{\boldsymbol{U}} = -\frac{\tau_i}{n c_{\boldsymbol{\tau}}(1+\tau_i)}(\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}^H)\boldsymbol{x_i}\boldsymbol{x_i}^H\boldsymbol{U},$$

$$(\boldsymbol{G}_{\boldsymbol{\tau}})_j = \begin{cases} 1 + \tau_i - \frac{1}{k}\boldsymbol{x_i}^H\boldsymbol{U}\boldsymbol{U}^H\boldsymbol{x_i} & \text{for} \quad j = i \\ 0 & \text{otherwise.} \end{cases}$$

Riemannian stochastic gradient descent

Input: Initial iterate $\bar{\theta}^{(1)} \in \overline{\mathcal{M}}_{p,k,n}$.

Output: Sequence of iterates $\{\bar{\theta}^{(t)}\}$.

$t = 1$

while *no convergence* **do**

Randomly draw a subset $A \subset \{\boldsymbol{x_i}\}_{i=1}^{n}$ and set

$\bar{\xi}^{(t)} = \sum_{x_i \in A} \operatorname{grad} \bar{L}_i(\bar{\theta}^{(t)})$

Compute a step size ν_t and set

$\bar{\theta}^{(t+1)} = \overline{R}_{\bar{\theta}^{(t)}}(-\nu_t \bar{\xi}^{(t)})$

$t = t + 1$

end

Algorithm 1: Riemannian stochastic gradient descent

Remark

Complexity of one iteration: $\mathcal{O}(mpk + pk^2)$ *where* $m = \#A$.

Riemannian geometry and clustering: application to a K-means++

Decoupled metric: geometry for distances

Definition

$\overline{\mathcal{M}}_{p,k,n}$ is endowed with the Riemannian metric defined by

$$\langle \bar{\xi}, \bar{\eta} \rangle_{\bar{\theta}}^{\overline{\mathcal{M}}_{p,k,n}} = \alpha \mathfrak{Re}(\mathrm{Tr}(\boldsymbol{\xi}_{\boldsymbol{U}}^H \boldsymbol{\eta}_{\boldsymbol{U}})) + \beta (\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\xi}_{\boldsymbol{\tau}})^T (\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\eta}_{\boldsymbol{\tau}}) \quad (14)$$

with $\alpha > 0$, $\beta > 0$.

From [2, 5] and properties of product manifolds:

Corollary (Distance)

The squared distance between θ_1 and θ_2 is

$$d^2_{\mathcal{M}_{p,k,n}}(\theta_1, \theta_2) = \alpha \left\|\Theta\right\|_2^2 + \beta \left\|\log(\boldsymbol{\tau}_1) - \log(\boldsymbol{\tau}_2)\right\|_2^2, \quad (15)$$

where $\boldsymbol{U}_1^H \boldsymbol{U}_2 \stackrel{SVD}{=} \boldsymbol{O}_1 \cos(\Theta) \boldsymbol{O}_2^H$.

Mean computation

The mean $c = \pi(\boldsymbol{U}, \boldsymbol{\tau})$ of the set of points $\{\theta_i = \pi(\boldsymbol{U}_i, \boldsymbol{\tau}_i)\}_{i=1}^{M}$ is obtained from the minimization of the variance,

$$c = \underset{\theta \in \mathcal{M}_{p,k,n}}{\arg\min} \frac{1}{2M} \sum_{i=1}^{M} d^2_{\mathcal{M}_{p,k,n}}(\theta, \theta_i). \tag{16}$$

Therefore, $\boldsymbol{\tau}$ is the geometric mean defined as

$$\boldsymbol{\tau} = \left(\prod_{\theta_i \in S_j}^{\odot} \boldsymbol{\tau}_i \right)^{\odot 1/m}, \tag{17}$$

where $\prod^{\odot}$ is the elementwise product.

A Riemannian gradient descent computes $\boldsymbol{U}$ (mean computation on the Grassmann manifold). Given $\boldsymbol{U}^{(t)}$, the iterate $\boldsymbol{U}^{(t+1)}$ is obtained with

$$\boldsymbol{U}^{(t+1)} = \exp^{\mathrm{Gr}_{p,k}}_{\boldsymbol{U}^{(t)}} \left(\frac{\nu_t}{M} \sum_{i=1}^{M} \log^{\mathrm{Gr}_{p,k}}_{\boldsymbol{U}^{(t)}} (\boldsymbol{U}_i) \right), \tag{18}$$

where ν_t is the step size.

Application in machine learning: K-means++ on $\mathcal{M}_{p,k,n}$

Clustering framework: the aim is to partition the descriptors $\{\theta_i\}_{i=1}^{M}$ in $S = \{S_1, S_2, \cdots, S_K\}$.

K-means++ on $\mathcal{M}_{p,k,n}$:
Initialization: recursively choose a new center θ_i with probability $\frac{D(\theta_i)^2}{\sum_{\theta_j} D(\theta_j)^2}$. $D(\theta_i)$ denotes the distance $d_{\mathcal{M}_{p,k,n}}$ from θ_i to the closest center among those already chosen.
Assignment step: $\forall i \in [\![1, M]\!]$ assign θ_i to the cluster S_j with the nearest c_j, $j \in [\![1, K]\!]$.
Update step: compute new centers c_j of clusters S_j, $\forall j \in [\![1, K]\!]$, using Riemannian means.

Application in machine learning: K-means++ on $\mathcal{M}_{p,k,n}$

K-means/*K-means++* optimize the within-cluster sum of squares:

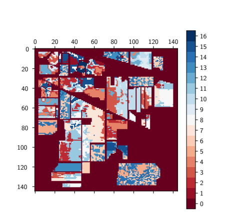
$$\phi(S) = \sum_{j=1}^{K} \sum_{\theta_i \in S_j} d^2_{\mathcal{M}_{p,k,n}}(c_j, \theta_i). \tag{19}$$

K-means++ on a Riemannian geometry is $\mathcal{O}(\log K)$ competitive with the optimal clustering:

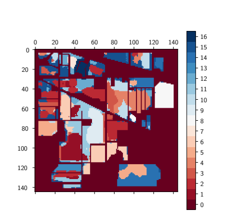
$$\mathbb{E}[\phi] \leq 8(\ln K + 2)\phi_{\mathrm{OPT}} \tag{20}$$

where ϕ_{OPT} is the minimum of ϕ and the expectation is taken with respect to the initialization procedure.

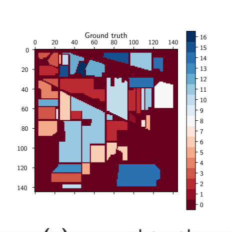
K-means++ on $\mathcal{M}_{p,k,n}$



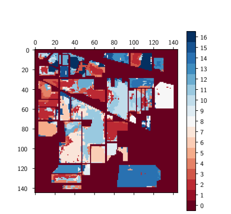
(a) K-means++ [3]: OA $= 31.2\%$



(b) K-means++ on "SCM/$\mathcal{H}_p^{++}$": OA $= 45.2\%$



(c) ground truth



(d) K-means++ on $\mathcal{M}_{p,k,n}$: OA $= 47.2\%$

Figure 7: *Indian Pines* [4] segmentation results achieved using different geometries/features.

Conclusion

To conclude, we presented:

1. the framework of optimization on matrix manifolds,
2. an estimation algorithm of the Probabilistic PCA from heteroscedastic signals model,
3. a *K-means++* on $\mathcal{M}_{p,k,n}$ with an application on hyperspectral data.

Questions ?

References

- P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton University Press, 2008. ISBN: 0691132984, 9780691132983.
- P.-A. Absil, R. Mahony, and R. Sepulchre. "Riemannian geometry of Grassmann manifolds with a view on algorithmic computation". In: *Acta Applicandae Mathematica* 80.2 (2004), pp. 199–220.
- D. Arthur and S. Vassilvitskii. "K-Means++: The Advantages of Careful Seeding". In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '07. New Orleans, Louisiana: Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035. ISBN: 9780898716245.

M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe. *220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3*. Sept. 2015. DOI: `doi:/10.4231/R7RX991C`. URL: `https://purr.purdue.edu/publications/1947/1`.

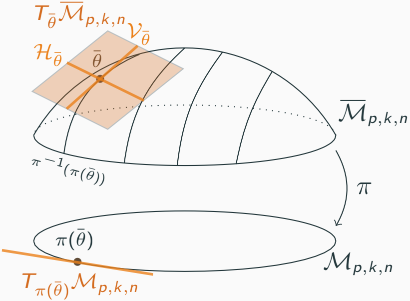
A. Edelman, T.A. Arias, and S. T. Smith. "The geometry of algorithms with orthogonality constraints". In: *SIAM journal on Matrix Analysis and Applications* 20.2 (1998), pp. 303–353.

Quotient manifold

Let $\overline{\mathcal{M}}$ be a smooth manifold and let $\sim$ define an equivalence relation over $\overline{\mathcal{M}}$. Every point $\bar{\theta} \in \overline{\mathcal{M}}$ belongs to an equivalence class

$$\pi(\bar{\theta}) = \{\bar{\theta}' \in \overline{\mathcal{M}} : \bar{\theta} \sim \bar{\theta}'\}.$$

Under conditions, the quotient space $\mathcal{M} = \overline{\mathcal{M}}/ \sim := \{\pi(\bar{\theta}) : \bar{\theta} \in \overline{\mathcal{M}}\}$, with the metric $\langle \cdot, \cdot \rangle_{\bar{\theta}}^{\overline{\mathcal{M}}}$, admits a unique Riemannian manifold structure. Then, the vertical space $\mathcal{V}_{\bar{\theta}}$ is defined as $\mathcal{V}_{\bar{\theta}} = T_{\bar{\theta}}\pi^{-1}(\pi(\bar{\theta}))$. The horizontal space $\mathcal{H}_{\bar{\theta}}$ is such that $T_{\bar{\theta}}\mathcal{M} = \mathcal{V}_{\bar{\theta}} \oplus \mathcal{H}_{\bar{\theta}}$.



Riemannian geometry of $\mathcal{M}_{p,k,n}$

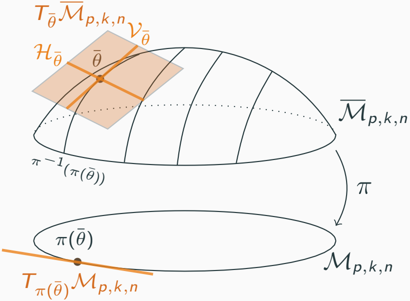
Vertical space of $\overline{\mathcal{M}}_{p,k,n}$ at $\bar{\theta} \in \overline{\mathcal{M}}_{p,k,n}$ is

$$\mathcal{V}_{\bar{\theta}} \triangleq T_{\bar{\theta}}\pi^{-1}(\pi(\bar{\theta})) = \{(\boldsymbol{UA}, 0) : \boldsymbol{A} \in \mathbb{C}^{k\times k},\ \boldsymbol{A}^H = -\boldsymbol{A}\}. \tag{21}$$

Horizontal space $\mathcal{H}_{\bar{\theta}}$: the orthogonal complement to $\mathcal{V}_{\bar{\theta}}$ in $T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k,n}$:

$$\mathcal{H}_{\bar{\theta}} = \{(\boldsymbol{\xi_U}, \boldsymbol{\xi_\tau}) \in \mathbb{C}^{p\times k} \times \mathbb{R}^n : \boldsymbol{U}^H\boldsymbol{\xi_U} = 0\}. \tag{22}$$

Hence, $\xi_{\bar{\theta}} \in \mathcal{H}_{\bar{\theta}}$ uniquely defines $\xi_\theta = \mathrm{D}\,\pi(\bar{\theta})[\xi_{\bar{\theta}}] \in T_{\pi(\bar{\theta})}\mathcal{M}_{p,k,n}$ and reciprocally.



Negative log-likelihood optimization: FIM vs decoupled metric

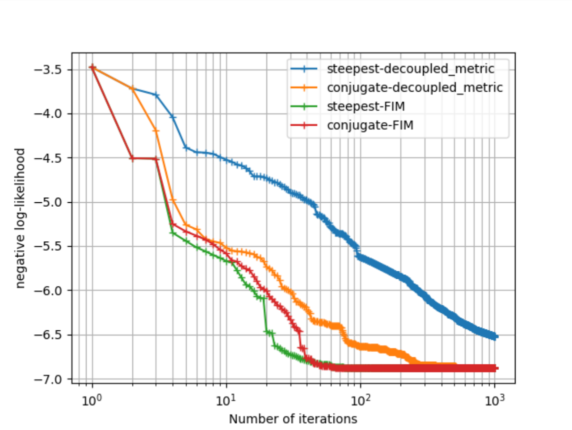


Figure 8: Riemannian gradient descent: Fisher information metric vs decoupled metric

Decoupled metric: geometry for distances

From [2, 5] and properties of product manifolds:

Corollary (Exponential mapping)

The exponential mapping on $\mathcal{M}_{p,k,n}$ is represented by

$$\exp_{\bar{\theta}}^{\overline{\mathcal{M}}_{p,k,n}}(\bar{\xi}) = \left(\exp_{\boldsymbol{U}}^{\mathrm{Gr}_{p,k}}(\boldsymbol{\xi}_{\boldsymbol{U}}), \exp_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\xi}_{\boldsymbol{\tau}})\right) \tag{23}$$

$\exp_{\boldsymbol{U}}^{\mathrm{Gr}_{p,k}}(\boldsymbol{\xi}_{\boldsymbol{U}}) = \boldsymbol{U}\boldsymbol{Y}\cos(\Sigma) + \boldsymbol{X}\sin(\Sigma)$, *with* $\boldsymbol{\xi}_{\boldsymbol{U}} \overset{SVD}{=} \boldsymbol{X}\Sigma\boldsymbol{Y}^T$,
$\exp_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\xi}_{\boldsymbol{\tau}}) = \boldsymbol{\tau} \odot \exp(\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\xi}_{\boldsymbol{\tau}})$.

Corollary (Logarithm mapping)

The logarithm map on $\mathcal{M}_{p,k,n}$ is represented by

$$\log_{\bar{\theta}_1}^{\overline{\mathcal{M}}_{p,k,n}}(\bar{\theta}_2) = \left(\log_{\boldsymbol{U}_1}^{\mathrm{Gr}_{p,k}}(\boldsymbol{U}_2), \log_{\boldsymbol{\tau}_1}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\tau}_2)\right) \tag{24}$$

$\log_{\boldsymbol{U}_1}^{\mathrm{Gr}_{p,k}}(\boldsymbol{U}_2) = \boldsymbol{X}\Theta\boldsymbol{Y}^H$ *where* $\boldsymbol{X}\Theta\boldsymbol{Y}^H$ *is computed with*
$(\boldsymbol{I}_p - \boldsymbol{U}_1\boldsymbol{U}_1^H)\boldsymbol{U}_2(\boldsymbol{U}_1^H\boldsymbol{U}_2)^{-1} \overset{SVD}{=} \boldsymbol{X}\tan(\Theta)\boldsymbol{Y}^H$,
$\log_{\boldsymbol{\tau}_1}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\tau}_2) = \boldsymbol{\tau}_1 \odot \log(\boldsymbol{\tau}_1^{\odot -1} \odot \boldsymbol{\tau}_2)$.