Riemannian geometry to robust estimation covariance matrices with application in machine learning

Antoine Collas

June 3, 2021

Supervisors: Jean-Philippe Ovarlez, Guillaume Ginolhac, Chengfang Ren, Arnaud Breloy.

Table of contents

Introduction

Context of the PhD

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

Problematics

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

- Change detection.
- Semantic segmentation.

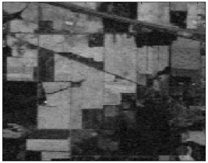


Figure 1: Raw image.

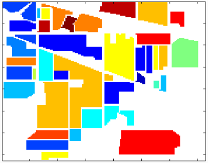


Figure 2: Segmented image. One color = one class (grass, woods, ...).

Objectives for parameter estimation

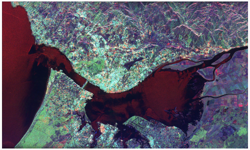


Figure 3: Example of a SAR image (from `nasa.gov`).

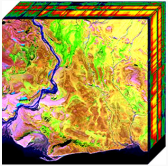


Figure 4: Example of a hyperspectral image (from `nasa.gov`).

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 handle strong noise of SAR images,
- "low-rank" estimators, i.e estimators that handle high dimension of hyperspectral images.

Riemannian geometry, optimization, and Intrinsic Cramér-Rao bounds

Riemannian geometry

A tool of interest for contrained parameters estimation is the Riemannian geometry. Briefly, 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*.

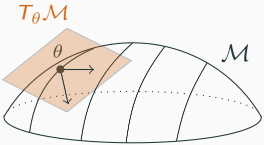


Figure 5: A manifold $\mathcal{M}$ with its tangent space $T_\theta\mathcal{M}$.

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

Then we define some important objects:

1. $\exp_\theta^{\mathcal{M}}(.) : T_\theta\mathcal{M} \to \mathcal{M}$, the Riemannian exponential mapping (extension of straight lines to Riemmanian manifolds),
2. $\log_\theta^{\mathcal{M}}(.) : \mathcal{M} \to T_\theta\mathcal{M}$, the Riemannian logarithm mapping,
3. $d(\cdot,\cdot)^{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$, the Riemannian distance.

Introduction to optimization on matrix manifolds

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

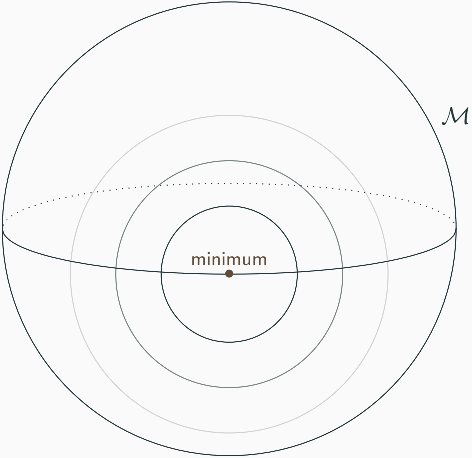
$$\min_{\theta \in \mathcal{M}} f(\theta)$$

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

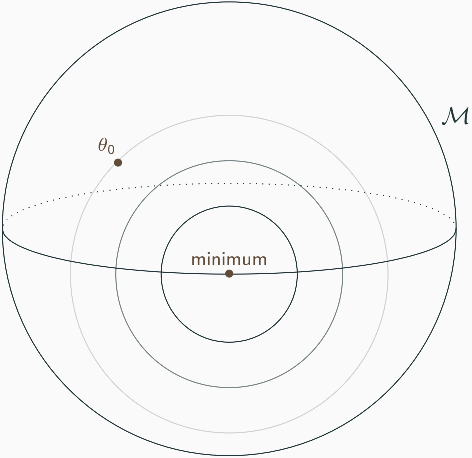
Examples of smooth manifolds $\mathcal{M}$:

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

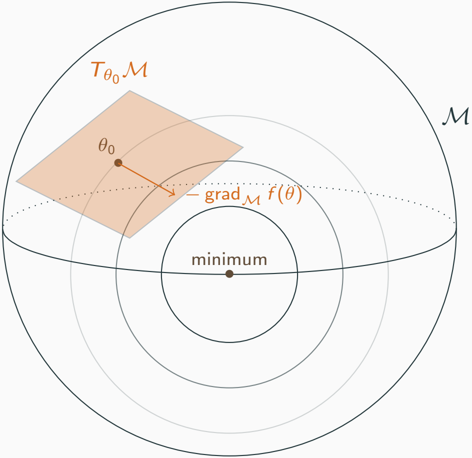
Introduction to optimization on matrix manifolds



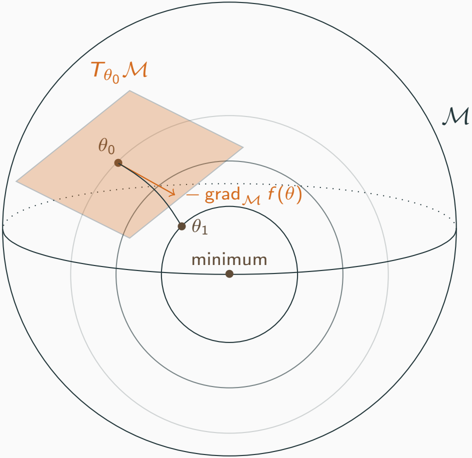
Introduction to optimization on matrix manifolds



Introduction to optimization on matrix manifolds



Introduction to optimization on matrix manifolds



Introduction to optimization on matrix manifolds

To minimize $f : \mathcal{M} \to \mathbb{R}$, where $\mathcal{M}$ is a Riemannian manifold, we need :

- $R_\theta^{\mathcal{M}}(\cdot)$: $T_\theta\mathcal{M} \to \mathcal{M}$: a retraction (generalization of the exponential mapping),
- $\operatorname{grad}_{\mathcal{M}} f(\theta)$: the Riemannian gradient of f at $\theta \in \mathcal{M}_{p,k,n}$, defined as,

$$\forall \xi \in T_\theta\mathcal{M}, \quad \mathrm{D}\, f(\theta)[\xi] = \langle \operatorname{grad}_{\mathcal{M}} f(\theta), \xi \rangle_\theta^{\mathcal{M}}.$$

Remark

Input: *Initialisation* $\theta_0 \in \mathcal{M}$, *step size* $\alpha > 0$, *number of iterations* K
Output: $\theta_K \in \mathcal{M}$
for $k = 0 : K - 1$ **do**
  $\theta_{k+1} = R_{\theta_k}^{\mathcal{M}}\left(-\alpha \operatorname{grad}_{\mathcal{M}} f(\theta_k)\right)$
end
Algorithm 1: Algorithm of steepest descent with a constant step size.

For a full review on this topic: Optimization algorithms on matrix manifolds [AMS08].

Intrinsic Cramér-Rao Bound (ICRB)

Let $\left(\mathcal{M}, \langle\cdot,\cdot\rangle_\theta^{\mathcal{M}}\right)$ be an n-dimensional Riemannian manifold with $\{e_\theta^q\}_{1\leq q\leq n}$ an orthonormal basis of $T_\theta\mathcal{M}$.

Let $\mathrm{X} \in \mathbb{C}^p$ be a random variable that admits a probability density function (pdf) $p(\mathrm{x}|\theta)$ which depends on a parameter $\theta \in \mathcal{M}$.

We can compute the Fisher Information Metric, $\forall \xi, \eta \in T_\theta\mathcal{M}$:

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

where L is the log-likelihood associated to the pdf.

Then, we can lower bound the variance of an unbiased estimator $\hat{\theta} \in \mathcal{M}$:

$$\mathbb{E}\left[d^2_{\mathcal{M}}(\theta, \hat{\theta})\right] \geq \mathrm{Tr}(\boldsymbol{F}_\theta^{-1}),$$

where the Fisher information matrix $\boldsymbol{F}_\theta$ is defined as $(\boldsymbol{F}_\theta)_{ql} = \langle e_\theta^q, e_\theta^l \rangle_\theta^{\mathrm{FIM}}$.

For a full review on this topic: Covariance, Subspace, and Intrinsic Cramér-Rao Bounds [Smi05].

A Tyler-type estimator of location and scatter

Compound Gaussian distribution - Data model

Let n data points $\boldsymbol{x_i} \in \mathbb{C}^p$ distributed according to the model:

$$\boldsymbol{x_i} \underset{d}{=} \boldsymbol{\mu} + \sqrt{\tau_i}\Sigma^{\frac{1}{2}}\boldsymbol{u_i} \tag{1}$$

where $\boldsymbol{\mu} \in \mathbb{C}^p$, $\boldsymbol{\tau} \in (\mathbb{R}_*^+)^n$, $\Sigma \in \mathcal{SH}_p^{++}$ and $\boldsymbol{u_i} \sim \mathbb{C}\mathcal{N}(0, \boldsymbol{I}_p)$, with

$$(\mathbb{R}_*^+)^n = \{\boldsymbol{\tau} \in \mathbb{R}^n, \boldsymbol{\tau}_i > 0\}, \tag{2}$$

$$\mathcal{SH}_p^{++} = \{\Sigma \in \mathcal{H}_p, \Sigma \succ 0, \det(\Sigma) = 1\}. \tag{3}$$

Thus, $\boldsymbol{x_i}$ follows a Compound Gaussian distribution, *i.e.*

$$\boldsymbol{x_i} \sim \mathbb{C}\mathcal{N}(\boldsymbol{\mu}, \tau_i\Sigma). \tag{4}$$

Definition

$$\mathcal{M}_{p,n} = \mathbb{C}^p \times (\mathbb{R}_*^+)^n \times \mathcal{SH}_p^{++} \tag{5}$$

Remark

The textures τ_i are assumed to be unknown and deterministic.

Data model - Log-likelihood

Hence, $\forall \theta = (\boldsymbol{\mu}, \boldsymbol{\tau}, \Sigma) \in \mathcal{M}_{p,n}$ the negative log-likelihood is

$$L(\theta) = \sum_{i=1}^{n} \left[\log\det\left(\tau_i \Sigma\right) + \frac{(\boldsymbol{x_i} - \boldsymbol{\mu})^H \Sigma^{-1} (\boldsymbol{x_i} - \boldsymbol{\mu})}{\tau_i} \right]. \qquad (6)$$

And the Maximum Likelihood Estimate satisfies

$$\begin{cases} \boldsymbol{\mu} = \left(\sum_{i=1}^{n} \frac{1}{\tau_i} \right)^{-1} \sum_{i=1}^{n} \frac{\boldsymbol{x_i}}{\tau_i} \\ \Sigma = \frac{1}{n} \sum_{i=1}^{n} \frac{(\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x_i} - \boldsymbol{\mu})^H}{\tau_i} \\ \tau_i = \frac{1}{p} (\boldsymbol{x_i} - \boldsymbol{\mu})^H \Sigma^{-1} (\boldsymbol{x_i} - \boldsymbol{\mu}). \end{cases} \qquad (7)$$

Remark

(7) *does not satisfy Maronna's conditions [Mar76], and the associated fixed-point iterations generally diverge in practice !*

$\mathcal{M}_{p,n}$: parameter manifold

The goal is to minimize the negative log-likelihood:

$$\hat{\theta} = \underset{\theta \in \mathcal{M}_{p,n}}{\arg\min}\, L(\theta). \tag{8}$$

The tangent space of $\mathcal{M}_{p,n}$ at θ is the product of the tangent spaces of $\mathbb{C}^p$, $(\mathbb{R}_*^+)^n$ and $\mathcal{SH}_p^{++}$ i.e,

$$T_\theta \mathcal{M}_{p,n} = \left\{ \xi \in \mathbb{C}^p \times \mathbb{R}^n \times \mathcal{H}_p : \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}_{\boldsymbol{\Sigma}}) = 0 \right\}, \tag{9}$$

where $\mathcal{H}_p$ is the Hermitian set.

Remark

$\mathcal{M}_{p,n}$ is a product manifold of sets which have well known Riemannian manifolds.

$\mathcal{M}_{p,n}$: Riemannian parameter manifold

Definition

Let $\xi, \eta \in T_\theta \mathcal{M}_{p,n}$, the Riemannian metric at θ is defined as,

$$\langle \xi, \eta \rangle_\theta^{\mathcal{M}_{p,n}} \quad = \quad \langle \boldsymbol{\xi}_{\boldsymbol{\mu}}, \boldsymbol{\eta}_{\boldsymbol{\mu}} \rangle_{\boldsymbol{\mu}}^{\mathbb{C}^p} \; + \; \langle \boldsymbol{\xi}_{\boldsymbol{\tau}}, \boldsymbol{\eta}_{\boldsymbol{\tau}} \rangle_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n} \; + \; \langle \boldsymbol{\xi}_\Sigma, \boldsymbol{\eta}_\Sigma \rangle_\Sigma^{\mathcal{H}_p^{++}}, \quad (10)$$

with

- $\langle \boldsymbol{\xi}_{\boldsymbol{\mu}}, \boldsymbol{\eta}_{\boldsymbol{\mu}} \rangle_{\boldsymbol{\mu}}^{\mathbb{C}^p} = \mathfrak{Re}\{\boldsymbol{\xi}_{\boldsymbol{\mu}}^H \boldsymbol{\eta}_{\boldsymbol{\mu}}\}$,
- $\langle \boldsymbol{\xi}_{\boldsymbol{\tau}}, \boldsymbol{\eta}_{\boldsymbol{\tau}} \rangle_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n} = (\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\xi}_{\boldsymbol{\tau}})^T (\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\eta}_{\boldsymbol{\tau}})$, where $\odot$ and $.^{\odot t}$ denote the elementwise product and power operators respectively,
- $\langle \boldsymbol{\xi}_\Sigma, \boldsymbol{\eta}_\Sigma \rangle_\Sigma^{\mathcal{H}_p^{++}} = \operatorname{Tr}\left(\Sigma^{-1} \boldsymbol{\xi}_\Sigma \Sigma^{-1} \boldsymbol{\eta}_\Sigma\right)$.

Remark

$\left(\mathcal{M}_{p,n}, \langle \cdot, \cdot \rangle_\cdot^{\mathcal{M}_{p,n}}\right)$ *is a Riemannian manifold and all its geometrical elements (exponential mapping, parallel transport, and distance) are derived from Riemannian geometries of* $\mathbb{C}^p$, $(\mathbb{R}_*^+)^n$, *and* $\mathcal{SH}_p^{++}$.

Riemannian optimization

Definition (Retraction)

$\forall \theta \in \mathcal{M}_{p,n}, \forall \xi \in T_\theta \mathcal{M}_{p,n}$,

$$R_\theta^{\mathcal{M}_{p,n}}(\xi) = \left(R_{\boldsymbol{\mu}}^{\mathbb{C}^p}(\boldsymbol{\xi}_{\boldsymbol{\mu}}), R_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\xi}_{\boldsymbol{\tau}}), R_{\Sigma}^{\mathcal{SH}_p^{++}}(\boldsymbol{\xi}_\Sigma) \right), \tag{11}$$

with

$R_{\boldsymbol{\mu}}^{\mathbb{C}^p}(\boldsymbol{\xi}_{\boldsymbol{\mu}}) = \boldsymbol{\mu} + \boldsymbol{\xi}_{\boldsymbol{\mu}}$,

$R_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\xi}_{\boldsymbol{\tau}}) = \boldsymbol{\tau} + \boldsymbol{\xi}_{\boldsymbol{\tau}} + \frac{1}{2}\boldsymbol{\tau}^{\odot -1}\boldsymbol{\xi}_{\boldsymbol{\tau}}^{\odot 2}$,

$R_{\Sigma}^{\mathcal{SH}_p^{++}}(\boldsymbol{\xi}_\Sigma) = \det\left(\Sigma + \boldsymbol{\xi}_\Sigma + \frac{1}{2}\boldsymbol{\xi}_\Sigma \Sigma^{-1}\boldsymbol{\xi}_\Sigma\right)^{-\frac{1}{p}} \left(\Sigma + \boldsymbol{\xi}_\Sigma + \frac{1}{2}\boldsymbol{\xi}_\Sigma \Sigma^{-1}\boldsymbol{\xi}_\Sigma\right)$.

Definition (Parallel transport)

$\forall \theta_1, \theta_2 \in \mathcal{M}_{p,n}, \forall \xi \in T_{\theta_1} \mathcal{M}_{p,n}$,

$$\mathcal{T}_{\theta_1,\theta_2}^{\mathcal{M}_{p,n}}(\xi) = \left(\mathcal{T}_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2}^{\mathbb{C}^p}(\boldsymbol{\xi}_{\boldsymbol{\mu}}), \mathcal{T}_{\boldsymbol{\tau}_1,\boldsymbol{\tau}_2}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\xi}_{\boldsymbol{\tau}}), \mathcal{T}_{\Sigma_1,\Sigma_2}^{\mathcal{SH}_p^{++}}(\boldsymbol{\xi}_\Sigma) \right), \tag{12}$$

with

$\mathcal{T}_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2}^{\mathbb{C}^p}(\boldsymbol{\xi}_{\boldsymbol{\mu}}) = \boldsymbol{\xi}_{\boldsymbol{\mu}}$,

$\mathcal{T}_{\boldsymbol{\tau}_1,\boldsymbol{\tau}_2}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\xi}_{\boldsymbol{\tau}}) = \boldsymbol{\tau}_2 \odot \boldsymbol{\tau}_1^{\odot -1} \odot \boldsymbol{\xi}_{\boldsymbol{\tau}}$,

$\mathcal{T}_{\Sigma_1,\Sigma_2}^{\mathcal{SH}_p^{++}}(\boldsymbol{\xi}_\Sigma) = \left(\Sigma_2\Sigma_1^{-1}\right)^{\frac{1}{2}} \boldsymbol{\xi}_\Sigma \left(\left(\Sigma_2\Sigma_1^{-1}\right)^{\frac{1}{2}} \right)^H$.

Riemannian optimization

Input : Initial iterate $\theta_1 \in \mathcal{M}_{p,n}$.
Output: Sequence of iterates $\{\theta_k\}$
$k := 1$;
$\xi_1 := -\operatorname{grad} L(\theta_1)$;
while *no convergence* **do**
  Compute a step size α_k (e.g see [AMS08, §4.2]) and set $\theta_{k+1} := R_{\theta_k}^{\mathcal{M}_{p,n}}(\alpha_k \xi_k)$;
  Compute β_{k+1} (e.g see [AMS08, §8.3]) and set $\xi_{k+1} := -\operatorname{grad} L(\theta_{k+1}) + \beta_{k+1} \mathcal{T}_{\theta_k,\theta_{k+1}}^{\mathcal{M}_{p,n}}(\xi_k)$;
  $k := k+1$;
end

Algorithm 2: Riemannian conjugate gradient [AMS08]

Remark

$\operatorname{grad} L(\theta_k)$ *is the Riemannian gradient of the negative log-likelihood.*

Numerical experiment

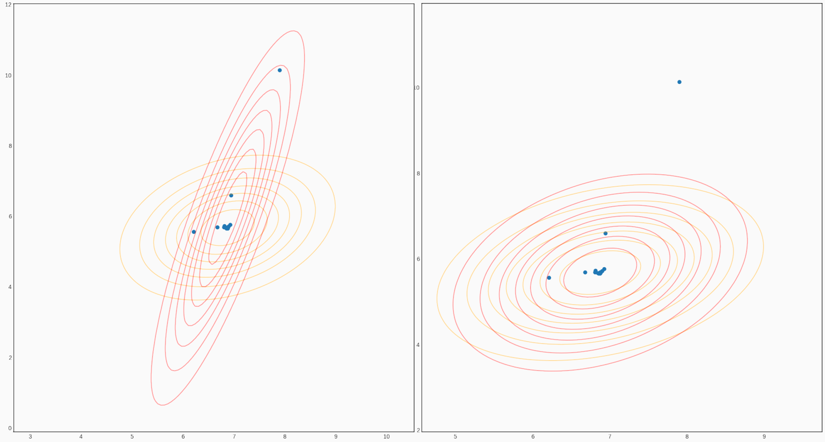


Figure 6: Example of a set of points generated with a heavy-tailed distribution with real probability density function (p.d.f.) in orange. Estimated p.d.f. are in red: Gaussian estimators on the left, our estimators on the right.

Numerical experiment

We compare the mean squared errors of different estimators on simulated data according to model (17).

1. Gaussian estimators: sample mean $\boldsymbol{\mu}^{\mathrm{G}}$ and SCM denoted Σ^{G}.
2. Two-step estimation: the sets $\{\boldsymbol{x_i}\}_{i=1}^n$ are centered with $\boldsymbol{\mu}^{\mathrm{G}}$ and then we estimate Σ using Tyler's M-estimator [Tyl87]. The estimator is denoted $\Sigma^{\mathrm{Ty},\boldsymbol{\mu}^{\mathrm{G}}}$.
3. Tyler's joint estimators of location and scatter matrix [Tyl87] denoted $\boldsymbol{\mu}^{\mathrm{Ty}}$ and Σ^{Ty}. It converges in practice unlike fixed-point equations of the MLE.
4. Tyler's M-estimator with location known [Tyl87]. The sets $\{\boldsymbol{x_i}\}_{i=1}^n$ are centered with $\boldsymbol{\mu}$ and then we estimate Σ. The estimator is denoted $\Sigma^{\mathrm{Ty},\boldsymbol{\mu}}$.
5. Our estimators $\boldsymbol{\mu}^{\mathrm{CG}}$ and Σ^{CG}: a Riemannian conjugate gradient to minimize (6) on $\mathcal{M}_{p,n}$ performed with the library *Pymanopt* [TKW16].

Numerical experiment

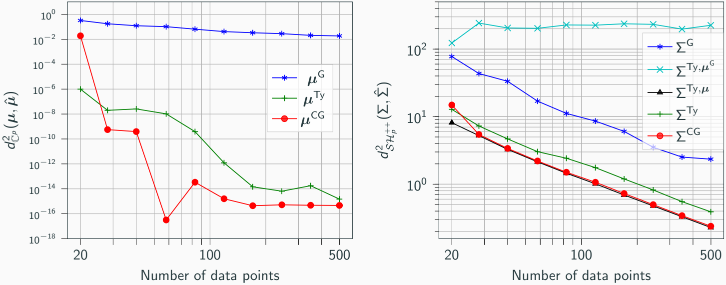


Figure 7: Mean squared errors over 200 simulated sets $\{\boldsymbol{x_i}\}_{i=1}^n$ ($p = 10$) with respect to the number n of samples for the considered estimators $\hat{\boldsymbol{\mu}} \in \{\boldsymbol{\mu}^{\mathrm{G}}, \boldsymbol{\mu}^{\mathrm{Ty}}, \boldsymbol{\mu}^{\mathrm{CG}}\}$ and $\hat{\Sigma} \in \{\Sigma^{\mathrm{G}}, \Sigma^{\mathrm{Ty},\boldsymbol{\mu}^{\mathrm{G}}}, \Sigma^{\mathrm{Ty},\boldsymbol{\mu}}, \Sigma^{\mathrm{Ty}}, \Sigma^{\mathrm{CG}}\}$.

Remark

$\boldsymbol{\mu}^{\mathrm{CG}}$ and Σ^{CG}, Riemannian Conjugate Gradient estimators, perform better than other estimators. For $n \geq 3p$, Σ^{CG} perform as good as Tyler's estimator with $\boldsymbol{\mu}$ known, $\Sigma^{\mathrm{Ty},\boldsymbol{\mu}}$, [Tyl87] !

A Geometry for Probabilistic PCA from heteroscedastic signals

Unitary matrices and complex Stiefel manifold

Definition

The unitary matrices are defined as:

$$\mathcal{U}_k = \{\boldsymbol{U} \in \mathbb{C}^{k\times k} : \boldsymbol{U}^H \boldsymbol{U} = \boldsymbol{I}_k\} \tag{13}$$

The complex Stiefel manifold is defined as:

$$\mathrm{St}_{p,k} = \{\boldsymbol{U} \in \mathbb{C}^{p\times k} : \boldsymbol{U}^H \boldsymbol{U} = \boldsymbol{I}_k\} \tag{14}$$

The complex Grassmann manifold is defined as:

$$\mathrm{Gr}_{p,k} = \mathrm{St}_{p,k}/\mathcal{U}_k = \{\pi(\boldsymbol{U}) :\ \boldsymbol{U} \in \mathrm{St}_{p,k}\}, \tag{15}$$

where π is such that

$$\pi(\boldsymbol{U}) = \{\boldsymbol{U}\boldsymbol{O} :\ \boldsymbol{O} \in \mathcal{U}_k\}. \tag{16}$$

Data model

Let n data points $\{\boldsymbol{x_i}\} \subset \mathbb{C}^p$ that are distributed as

$$\boldsymbol{x_i} \underset{d}{=} \sqrt{\tau_i}\boldsymbol{U}\boldsymbol{g_i} + \boldsymbol{n_i} \tag{17}$$

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

Hence,

$$\boldsymbol{x_i} \sim \mathbb{C}\mathcal{N}\left(0, \bar{\psi}_i(\bar{\theta}) \triangleq \boldsymbol{I}_p + \tau_i \boldsymbol{U}\boldsymbol{U}^H\right) \tag{18}$$

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 $\boldsymbol{O} \in \mathcal{U}_k$, $\bar{\psi}_i(\boldsymbol{U}\boldsymbol{O}, \boldsymbol{\tau}) = \bar{\psi}_i(\boldsymbol{U}, \boldsymbol{\tau})$.

Hence, each $\bar{\psi}_i$ induces a function ψ_i on the product manifold $\mathcal{M}_{p,k,n} \triangleq \mathrm{Gr}_{p,k} \times (\mathbb{R}_*^+)^n$ such that $\psi_i(\pi(\boldsymbol{U}), \boldsymbol{\tau}) = \bar{\psi}_i(\boldsymbol{U}, \boldsymbol{\tau})$.

The negative log-likelihood function, defined on $\mathcal{M}_{p,k,n}$, of our model is

$$L(\theta) = \sum_i \log\det \psi_i(\theta) + \boldsymbol{x_i}^H (\psi_i(\theta))^{-1} \boldsymbol{x_i}. \tag{19}$$

Riemannian geometry of manifold $\overline{\mathcal{M}}_{p,k,n} = \mathbf{St}_{p,k} \times (\mathbb{R}_*^+)^n$

Since $\overline{\mathcal{M}}_{p,k,n} = \mathrm{St}_{p,k} \times (\mathbb{R}_*^+)^n$, the tangent space of $\overline{\mathcal{M}}_{p,k,n}$ at $\bar{\theta} \in \overline{\mathcal{M}}_{p,k,n}$ is

$$T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k,n} = T_{\boldsymbol{U}}\mathrm{St}_{p,k} \times T_{\boldsymbol{\tau}}(\mathbb{R}_*^+)^n \tag{20}$$

where $T_{\boldsymbol{U}}\mathrm{St}_{p,k} = \{\boldsymbol{\xi}_{\boldsymbol{U}} \in \mathbb{C}^{p\times k} : \boldsymbol{U}^H\boldsymbol{\xi}_{\boldsymbol{U}} + \boldsymbol{\xi}_{\boldsymbol{U}}^H\boldsymbol{U} = 0\}$, and $T_{\boldsymbol{\tau}}(\mathbb{R}_*^+)^n$ is identified to $\mathbb{R}^n$.

Definition

$\overline{\mathcal{M}}_{p,k,n}$ is endowed with the Riemannian metric defined, for $\bar{\theta} \in \overline{\mathcal{M}}_{p,k,n}$, $\bar{\xi} = (\boldsymbol{\xi}_{\boldsymbol{U}}, \boldsymbol{\xi}_{\boldsymbol{\tau}}), \bar{\eta} = (\boldsymbol{\eta}_{\boldsymbol{U}}, \boldsymbol{\eta}_{\boldsymbol{\tau}}) \in T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k,n}$, by

$$\langle \bar{\xi}, \bar{\eta} \rangle_{\bar{\theta}}^{\overline{\mathcal{M}}_{p,k,n}} = \alpha \langle \boldsymbol{\xi}_{\boldsymbol{U}}, \boldsymbol{\eta}_{\boldsymbol{U}} \rangle_{\boldsymbol{U}}^{\mathrm{St}_{p,k}} + \beta \langle \boldsymbol{\xi}_{\boldsymbol{\tau}}, \boldsymbol{\eta}_{\boldsymbol{\tau}} \rangle_{\boldsymbol{\tau}}^{(\mathbb{R}^{++})^n} \tag{21}$$

where

$\langle \boldsymbol{\xi}_{\boldsymbol{U}}, \boldsymbol{\eta}_{\boldsymbol{U}} \rangle_{\boldsymbol{U}}^{\mathrm{St}_{p,k}} = \mathfrak{Re}(\mathrm{Tr}(\boldsymbol{\xi}_{\boldsymbol{U}}^H \boldsymbol{\eta}_{\boldsymbol{U}}))$,

$\langle \boldsymbol{\xi}_{\boldsymbol{\tau}}, \boldsymbol{\eta}_{\boldsymbol{\tau}} \rangle_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n} = (\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\xi}_{\boldsymbol{\tau}})^T (\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\eta}_{\boldsymbol{\tau}})$,

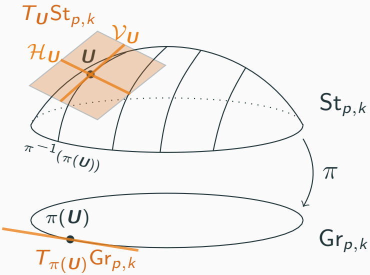
$\alpha > 0$, $\beta > 0$.

Riemannian geometry of manifold $\mathcal{M}_{p,k,n} = \mathbf{Gr}_{p,k} \times (\mathbb{R}_*^+)^n$

From [EAS98; AMS04], the tangent space $T_{\pi(\boldsymbol{U})}\mathrm{Gr}_{p,k}$ is represented by a subspace of $T_{\boldsymbol{U}}\mathrm{St}_{p,k}$: the horizontal space, which is

$$\mathcal{H}_{\boldsymbol{U}} = \{\boldsymbol{\xi}_{\boldsymbol{U}} \in \mathbb{C}^{p\times k} : \ \boldsymbol{U}^H \boldsymbol{\xi}_{\boldsymbol{U}} = 0\}. \quad (22)$$

A vector $\boldsymbol{\xi}_{\boldsymbol{U}} \in \mathcal{H}_{\boldsymbol{U}}$ uniquely defines the tangent vector $\xi_{\pi(\boldsymbol{U})} = \mathrm{D}\,\pi(\boldsymbol{U})[\boldsymbol{\xi}_{\boldsymbol{U}}] \in T_{\pi(\boldsymbol{U})}\mathrm{Gr}_{p,k}$.



$\theta = (\pi(\boldsymbol{U}), \boldsymbol{\tau}) \in \mathcal{M}_{p,k,n}$ is represented by $\bar{\theta} = (\boldsymbol{U}, \boldsymbol{\tau}) \in \overline{\mathcal{M}}_{p,k,n}$.

Corollary

The tangent space $T_\theta \mathcal{M}_{p,k,n}$ is represented by $\mathcal{H}_{\boldsymbol{U}} \times T_{\boldsymbol{\tau}}(\mathbb{R}_*^+)^n$ which is included in $T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k,n} = T_{\boldsymbol{U}}\mathrm{St}_{p,k} \times T_{\boldsymbol{\tau}}(\mathbb{R}_*^+)^n$.

Riemannian geometry of manifold $\mathcal{M}_{p,k,n} = \mathbf{Gr}_{p,k} \times (\mathbb{R}_*^+)^n$

From [AMS04; EAS98] and properties of product manifolds:

Corollary (Exponential mapping)

For $\theta \in \mathcal{M}_{p,k,n}, \xi \in T_\theta \mathcal{M}_{p,k,n}$, the exponential mapping is,

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

$\exp_{\pi(\boldsymbol{U})}^{\mathrm{Gr}_{p,k}}(\xi_{\pi(\boldsymbol{U})}) = \pi(\boldsymbol{U}\boldsymbol{Y}\cos(\Sigma) + \boldsymbol{X}\sin(\Sigma))$, *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)

Let $\theta_1 = (\pi(\boldsymbol{U}_1), \boldsymbol{\tau}_1), \theta_2 = (\pi(\boldsymbol{U}_2), \boldsymbol{\tau}_2) \in \mathcal{M}_{p,k,n}$, the logarithm map is,

$$\log_{\theta_1}^{\mathcal{M}_{p,k,n}}(\theta_2) = \left(\log_{\pi(\boldsymbol{U}_1)}^{\mathrm{Gr}_{p,k}}(\pi(\boldsymbol{U}_2)), \log_{\boldsymbol{\tau}_1}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\tau}_2) \right) \tag{24}$$

$\log_{\pi(\boldsymbol{U}_1)}^{\mathrm{Gr}_{p,k}}(\pi(\boldsymbol{U}_2)) = \mathrm{D}\,\pi(\boldsymbol{U}_1)[\boldsymbol{X}\Theta\boldsymbol{Y}^H]$ *where* $\boldsymbol{X}\Theta\boldsymbol{Y}^H \in \mathcal{H}_{\boldsymbol{U}_1}$ *is defined through the SVD* $(\boldsymbol{I}_p - \boldsymbol{U}_1\boldsymbol{U}_1^H)\boldsymbol{U}_2(\boldsymbol{U}_1^H\boldsymbol{U}_2)^{-1} \overset{SVD}{=} \boldsymbol{X}\tan(\Theta)\boldsymbol{Y}^H$,
$\log_{\boldsymbol{\tau}_1}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\tau}_2) = \boldsymbol{\tau}_1 \odot \log(\boldsymbol{\tau}_1^{\odot -1} \odot \boldsymbol{\tau}_2)$.

Riemannian geometry of manifold $\mathcal{M}_{p,k,n} = \mathbf{Gr}_{p,k} \times (\mathbb{R}_*^+)^n$

From [AMS04; EAS98] and properties of product manifolds:

Corollary (Distance)

Finally, the corresponding distance between θ_1 and θ_2 is

$$d^2_{\mathcal{M}_{p,k,n}}(\theta_1, \theta_2) = \alpha d^2_{\mathrm{Gr}_{p,k}}(\pi(\boldsymbol{U}_1), \pi(\boldsymbol{U}_2)) + \beta d^2_{(\mathbb{R}_*^+)^n}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2), \tag{25}$$

where

$d^2_{\mathrm{Gr}_{p,k}}(\pi(\boldsymbol{U}_1), \pi(\boldsymbol{U}_2)) = \|\Theta\|_2^2$ *where* $\boldsymbol{U}_1^H \boldsymbol{U}_2 \overset{SVD}{=} \boldsymbol{O}_1 \cos(\Theta) \boldsymbol{O}_2^H$,

$d^2_{(\mathbb{R}_*^+)^n}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \|\log(\boldsymbol{\tau}_1) - \log(\boldsymbol{\tau}_2)\|_2^2$.

Intrinsic Cramér-Rao Bound

From [Smi05]:

4 steps to get an Intrinsic Cramér-Rao Bound:

1. compute the Fisher Information Metric:

$$\langle \bar{\xi}, \bar{\eta} \rangle_{\bar{\theta}}^{\mathrm{FIM}} = \mathbb{E}[\mathrm{D}\, \bar{L}(\bar{\theta})[\bar{\xi}]\, \mathrm{D}\, \bar{L}(\bar{\theta})[\bar{\eta}]], \tag{26}$$

2. provide an orthonormal basis $\{e_{\bar{\theta}}^{q}\}_{1 \leq q \leq 2(p-k)k+n}$ of $T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k,n}$,
3. compute the Fisher information matrix $\boldsymbol{F}_{\bar{\theta}}$. The ql^{th} element of $\boldsymbol{F}_{\bar{\theta}}$ is defined as

$$(\boldsymbol{F}_{\bar{\theta}})_{ql} = \langle e_{\bar{\theta}}^{q}, e_{\bar{\theta}}^{l} \rangle_{\bar{\theta}}^{\mathrm{FIM}}, \tag{27}$$

4. lower bound the variance:

$$\mathbb{E}\left[d_{\mathcal{M}_{p,k,n}}^{2}(\theta, \hat{\theta}) \right] \geq \mathrm{Tr}(\boldsymbol{F}_{\bar{\theta}}^{-1}). \tag{28}$$

Fisher information metric and orthonormal basis of $T_{\bar\theta}\overline{\mathcal{M}}_{p,k,n}$

Proposition (Fisher information metric)

Given $\theta = (\pi(\boldsymbol{U}), \boldsymbol{\tau}) \in \mathcal{M}_{p,k,n}$, $\bar\xi = (\boldsymbol{\xi_U}, \boldsymbol{\xi_\tau}) \in T_{\bar\theta}\overline{\mathcal{M}}_{p,k,n}$ *and* $\bar\eta = (\boldsymbol{\eta_U}, \boldsymbol{\eta_\tau}) \in T_{\bar\theta}\overline{\mathcal{M}}_{p,k,n}$ *the Fisher information metric on* $\overline{\mathcal{M}}_{p,k,n}$ *corresponding to the log-likelihood is*

$$\begin{aligned}\langle\bar\xi,\bar\eta\rangle_{\bar\theta}^{\mathrm{FIM}} &= 2nc_{\boldsymbol{\tau}}\mathfrak{Re}(\mathrm{Tr}(\boldsymbol{\xi_U}^H\boldsymbol{\eta_U})) \\ &+ k(\boldsymbol{\xi_\tau}\odot(1+\boldsymbol{\tau})^{\odot -1})^T(\boldsymbol{\eta_\tau}\odot(1+\boldsymbol{\tau})^{\odot -1})\end{aligned} \tag{29}$$

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

Proposition (Orthonormal basis of $T_{\bar\theta}\overline{\mathcal{M}}_{p,k,n}$)

Given $\bar\theta \in \overline{\mathcal{M}}_{p,k,n}$*, an orthonormal basis of* $T_{\bar\theta}\overline{\mathcal{M}}_{p,k,n}$ *is*

$$\left\{\{(\alpha^{-\frac{1}{2}}\boldsymbol{U}_\perp\boldsymbol{K_{ij}},0),(\alpha^{-\frac{1}{2}}i\boldsymbol{U}_\perp\boldsymbol{K_{ij}},0)\}_{\substack{1\leq i\leq p-k\\1\leq j\leq k}},\{(0,\beta^{-\frac{1}{2}}\tau_i\boldsymbol{e_i})\}_{1\leq i\leq n}\right\} \tag{30}$$

where $\boldsymbol{U}_\perp \in \mathrm{St}_{p,p-k}$ *is such that* $\boldsymbol{U}^H\boldsymbol{U}_\perp = 0$, $\boldsymbol{K_{ij}} \in \mathbb{R}^{(p-k)\times k}$ *with* ij^{th} *element is* 1*, zeros elsewhere and* $\boldsymbol{e_i} \in \mathbb{R}^n$ *: its* i^{th} *element is* 1*, zeros elsewhere.*

Fisher information matrix $\boldsymbol{F}_{\bar{\theta}}$

Proposition (Fisher information matrix)

The Fisher information matrix $\boldsymbol{F}_{\bar{\theta}}$ on $\overline{\mathcal{M}}_{p,k,n}$ admits the structure

$$\boldsymbol{F}_{\bar{\theta}} = \begin{pmatrix} \boldsymbol{F}_{\boldsymbol{U}} & 0 \\ 0 & \boldsymbol{F}_{\boldsymbol{\tau}} \end{pmatrix} \tag{31}$$

with

$\boldsymbol{F}_{\boldsymbol{U}} = 2\,\alpha^{-1}\,n\,c_{\boldsymbol{\tau}}\,\boldsymbol{I}_{2(p-k)k}$, *and*

$\boldsymbol{F}_{\boldsymbol{\tau}} = \beta^{-1}\,k\,\boldsymbol{diag}\left(\boldsymbol{\tau}^{\odot 2}\odot(1+\boldsymbol{\tau})^{\odot -2}\right)$,

where $\boldsymbol{diag}(\cdot)$ returns the diagonal matrix formed with the elements of its argument.

ICRB

Proposition (Intrinsic Cramér-Rao bound on $\mathrm{Gr}_{p,k}$)

Given $\hat{\boldsymbol{U}} \in \mathrm{St}_{p,k}$ an estimation of $\boldsymbol{U} \in \mathrm{St}_{p,k}$, the Intrinsic Cramér-Rao bound which lower bounds the subspace estimation error is

$$\mathbb{E}\left[d^2_{\mathrm{Gr}_{p,k}}(\pi(\boldsymbol{U}), \pi(\hat{\boldsymbol{U}}))\right] \geq \frac{(p-k)k}{n c_{\boldsymbol{\tau}}} \triangleq \mathrm{CRB}_{\boldsymbol{U}} \tag{32}$$

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

Proposition (Intrinsic Cramér-Rao bound on $(\mathbb{R}_*^+)^n$)

Given $\hat{\boldsymbol{\tau}} \in (\mathbb{R}_*^+)^n$ an estimation of $\boldsymbol{\tau} \in (\mathbb{R}_*^+)^n$, the Intrinsic Cramér-Rao bound is

$$\mathbb{E}\left[d^2_{(\mathbb{R}_*^+)^n}(\boldsymbol{\tau}, \hat{\boldsymbol{\tau}})\right] \geq \frac{1}{k}\sum_{i=1}^{n} \frac{(1+\tau_i)^2}{\tau_i^2} \triangleq \mathrm{CRB}_{\boldsymbol{\tau}} \tag{33}$$

Numerical experiment

We generate sets $\{\boldsymbol{x_i}\}_{i=1}^n$, with,

$$\boldsymbol{x_i} \sim \mathbb{C}\mathcal{N}(0, \boldsymbol{I}_p + \tau_i \boldsymbol{U}\boldsymbol{U}^H) \tag{34}$$

and

$$\tau_i \sim \text{Log-normal}\left(\frac{-s^2}{2}, s^2\right) \tag{35}$$

$$\tau_i \leftarrow \text{SNR} \times \tau_i. \tag{36}$$

Here are the considered estimators in the simulations:

1. SCM: the k first principal eigenvectors of the SCM of $\{\boldsymbol{x_i}\}_{i=1}^n$ are concatenated to get $\boldsymbol{U}^{\text{SCM}}$.
2. BCD: the MLE estimate is done using a block coordinate descent (BCD) on $\{\boldsymbol{x_i}\}_{i=1}^n$ [Bre+13]. The estimators are denoted $\boldsymbol{U}^{\text{BCD}}$ and $\boldsymbol{\tau}^{\text{BCD}}$.
3. RO: Riemannian optimization (RO) is performed on $\{\boldsymbol{x_i}\}_{i=1}^n$ using a Riemannian conjugate gradient. Pymanopt library [TKW16] (builds upon the Manopt library [Bou+14]) achieves this optimization. The estimators are denoted $\boldsymbol{U}^{\text{RO}}$ and $\boldsymbol{\tau}^{\text{RO}}$.

Numerical experiment

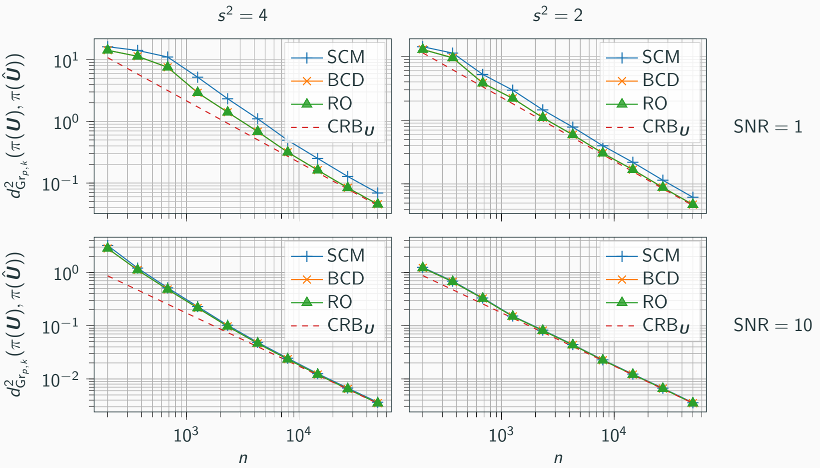


Figure 8: MSE over $N = 100$ simulated sets $\{\boldsymbol{x_i}\}$ ($p = 100$ and $k = 20$) with respect to the number of samples n for the three considered estimators. The textures are generated with $s^2 = 4$ (left part), $s^2 = 2$ (right part), SNR = 1 (upper part), SNR = 10 (lower part).

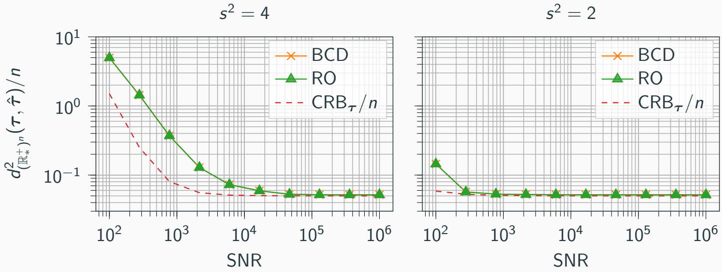


Figure 9: MSE over $N = 100$ simulated sets $\{\boldsymbol{x_i}\}$ ($n = 10^4$, $p = 100$ and $k = 20$) with respect to the SNR for the BCD and RO estimators. The textures are generated with $s^2 = 4$ (left) and $s^2 = 2$ (right).

Application in machine learning: mean computation of $\{\theta_i\}$

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

Let $m = \#S_j$, the variance V_j of S_j at $\theta \in \mathcal{M}_{p,k,n}$ is defined as,

$$V_j(\theta) = \frac{1}{m} \sum_{\theta_i \in S_j} d^2_{\mathcal{M}_{p,k,n}}(\theta, \theta_i). \tag{37}$$

The mean c_j of the set of points S_j is obtained from the minimization of the variance,

$$c_j = \underset{\theta \in \mathcal{M}_{p,k,n}}{\arg\min} \frac{1}{2} V_j(\theta). \tag{38}$$

As in the case of [Kar77], the gradient is

$$\operatorname{grad} V_j(\theta) = -\frac{1}{m} \sum_{\theta_i \in S_j} \log_{\theta}^{\mathcal{M}_{p,k,n}}(\theta_i). \tag{39}$$

Consequently, $c_j = (\pi(\boldsymbol{U}), \boldsymbol{\tau})$ satisfies,

$$\sum_{\theta_i \in S_j} \log_{c_j}^{\mathcal{M}_{p,k,n}}(\theta_i) = 0. \tag{40}$$

Application in machine learning: mean computation of $\{\theta_i\}$

Hence,

$$\begin{cases} \sum_{\theta_i \in S_j} \log_{\pi(\boldsymbol{U})}^{\mathrm{Gr}_{p,k}}(\pi(\boldsymbol{U}_i)) = 0, \\ \sum_{\theta_i \in S_j} \log_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\tau}_i) = 0. \end{cases} \tag{41}$$

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

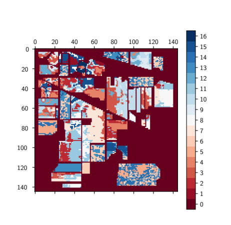
$$\boldsymbol{\tau} = \left(\prod_{\theta_i \in S_j}^{\odot} \boldsymbol{\tau}_i \right)^{\odot 1/m}, \tag{42}$$

where $\prod^{\odot}$ is the elementwise product. To estimate the mean of subspaces $\{\pi(\boldsymbol{U}_i)\}$ we use a Riemannian gradient descent. Given $\pi(\boldsymbol{U}^{(t)}) \in \mathrm{Gr}_{p,k}$, the iterate $\pi(\boldsymbol{U}^{(t+1)}) \in \mathrm{Gr}_{p,k}$ is obtained with

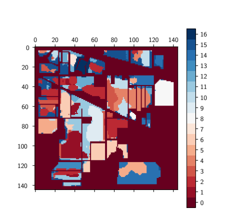
$$\pi(\boldsymbol{U}^{(t+1)}) = \exp_{\pi(\boldsymbol{U}^{(t)})}^{\mathrm{Gr}_{p,k}} \left(\frac{\nu_t}{m} \sum_{\theta_i \in S_j} \log_{\pi(\boldsymbol{U}^{(t)})}^{\mathrm{Gr}_{p,k}}(\pi(\boldsymbol{U}_i)) \right), \tag{43}$$

where ν_t is the step size.

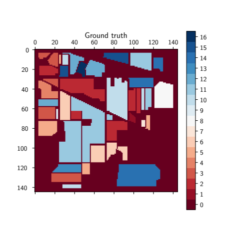
K-means++ on $\mathcal{M}_{p,k,n}$



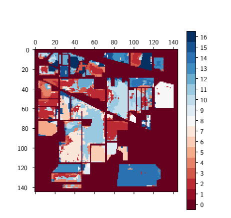
(a) K-means++ [AV07]: OA = 31.2%



(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 10: *Indian Pines* [BBL15] segmentation results achieved using different geometries/features.

Conclusion

To conclude, we presented:

1. the notion of Riemannian geometry,
2. the framework of optimization on matrix manifolds,
3. the concept of Intrinsic Cramér-Rao bound,
4. an algorithm of clustering on a Riemannian manifold with 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. "Riemannian geometry of Grassmann manifolds with a view on algorithmic computation". In: *Acta Applicandae Mathematica* 80.2 (2004), pp. 199–220.

P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton University Press, 2008. ISBN: 0691132984, 9780691132983.

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.

N. Boumal et al. "Manopt, a Matlab Toolbox for Optimization on Manifolds". In: *Journal of Machine Learning Research* 15 (2014), pp. 1455–1459.

A. Breloy et al. "Maximum likelihood estimation of clutter subspace in non homogeneous noise context". In: *21st European Signal Processing Conference (EUSIPCO 2013)*. Sept. 2013, pp. 1–5.

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.

H. Karcher. "Riemannian center of mass and mollifier smoothing". In: *Communications on Pure and Applied Mathematics* 30.5 (1977), pp. 509–541. DOI: https://doi.org/10.1002/cpa.3160300502. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160300502. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160300502.

R. A. Maronna. "Robust M-Estimators of Multivariate Location and Scatter". In: *Ann. Statist.* 4.1 (Jan. 1976), pp. 51–67. DOI: 10.1214/aos/1176343347. URL: https://doi.org/10.1214/aos/1176343347.

S. Smith. "Covariance, Subspace, and Intrinsic Cramér-Rao Bounds". In: *Signal Processing, IEEE Transactions on* 53 (June 2005), pp. 1610–1630. DOI: 10.1109/TSP.2005.845428.

J. Townsend, N. Koep, and S. Weichwald. "Pymanopt: A Python Toolbox for Optimization on Manifolds Using Automatic Differentiation". In: *J. Mach. Learn. Res.* 17.1 (Jan. 2016), pp. 4755–4759. ISSN: 1532-4435.

D. E. Tyler. "A Distribution-Free M-Estimator of Multivariate Scatter". In: *Ann. Statist.* 15.1 (Mar. 1987), pp. 234–251. DOI: 10.1214/aos/1176350263. URL: https://doi.org/10.1214/aos/1176350263.

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 $\langle \cdot, \cdot \rangle_{\theta}^{\overline{\mathcal{M}}}$, 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 $T_{\bar{\theta}}\mathcal{M} = \mathcal{V}_{\bar{\theta}} \oplus \mathcal{H}_{\bar{\theta}}$.