Riemannian geometry in elliptical distributions

Arnaud Breloy, SLSIP Workshop, Rüdesheim, October 7th 2021

Université Paris Nanterre

Some references

"*Intrinsic Cramér–Rao bounds for scatter and shape matrices estimation in CES distributions*," SPL, 2018.
Arnaud Breloy, Guillaume Ginolhac, Alexandre Renaux, Florent Bouchard

"*A Riemannian Framework for Low-Rank Structured Elliptical Models*," TSP, 2021.
Florent Bouchard, Arnaud Breloy, Guillaume Ginolhac, Alexandre Renaux, Frederic Pascal

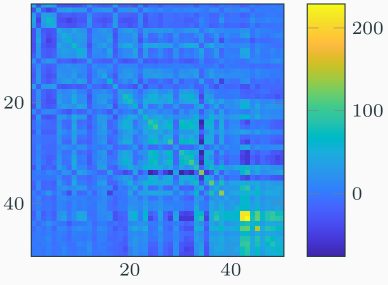
"*A Tyler-Type Estimator of Location and Scatter Leveraging Riemannian Optimization*," ICASSP 2021.
Antoine Collas, Florent Bouchard, Arnaud Breloy, Guillaume Ginolhac, Chengfang Ren, Jean-Philippe Ovarlez

"*Probabilistic PCA from Heteroscedastic Signals: Geometric Framework and Application to Clustering*"
Antoine Collas, Florent Bouchard, Arnaud Breloy, Guillaume Ginolhac, Chengfang Ren, Jean-Philippe Ovarlez

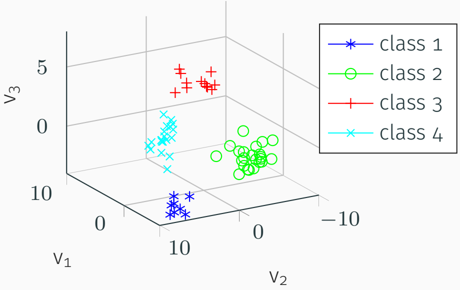
Motivations

Represent or analyze the data $\mathbf{x}$ through some parameter $\boldsymbol{\theta}$

Example with $p \simeq 7k$ genes of $n = 63$ patients with $k = 4$ classes [Khan2001] represented by



Covariance of 50 selected genes



3 principal components

Statistical approach

"Assume $\mathbf{x} \sim f(\mathbf{x}, \boldsymbol{\theta})$, then do stuff"

- **Design** a meaningful pdf f and parameter $\boldsymbol{\theta}$
- **Analyze** model properties, performance bounds...
- **Solve** related optimization problems (MLEs, barycenters...)
- **Apply** the results to a task

Today's talk: What can Riemannian geometry bring to these steps?

Outline

- ✗ **Design**
  - Examples of f and $\boldsymbol{\theta}$ from elliptical distributions
  - Remark that $\boldsymbol{\theta} \in \mathcal{M} \implies$ pretext to define Riemannian tools
- **Analyze**
  - Intrinsic Cramér-Rao bounds
  - 2 examples of interesting inequalities
- **Solve**
  - Riemannian optimization and geodesic convexity
  - 2 examples where numerical stability is improved
- **Apply**
  - Clustering with Riemannian distances

Elliptical distributions

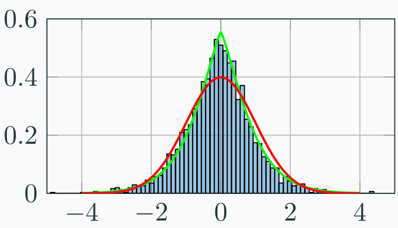
Complex elliptically symmetric distributions (CES)

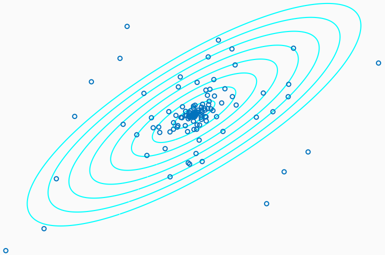
$\mathbf{x} \sim \mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ if it has for pdf

$$f(\mathbf{x}) \propto |\boldsymbol{\Sigma}|^{-1} g\left((\mathbf{x}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right)$$

Scaled Gaussian distributions (CG)

$\mathbf{x}_i \sim \mathcal{CN}(\boldsymbol{\mu}, \tau_i \boldsymbol{\Sigma})$ where τ_i is unknown deterministic





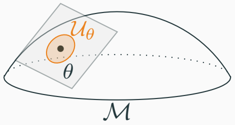
Structured parameter space as a manifold

Generally, the distribution **parameter space**, *e.g.*

- **Covariance matrices**: $\boldsymbol{\Sigma} \in \mathcal{H}_p^{++}$
- **Product spaces**: $\{\{\tau_i\}_{i=1}^n,\ \boldsymbol{\mu},\ \boldsymbol{\Sigma}\} \in (\mathbb{R}^+)^n \times \mathbb{C}^p \times \mathcal{H}_p^{++}$

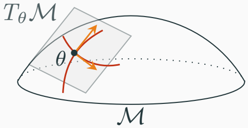
turn out to be a **manifold** $\mathcal{M}$ (locally diffeomorphic to $\mathbb{R}^d$, with $\dim(\mathcal{M}) = d$)

$$\forall \theta \in \mathcal{M},\ \exists \mathcal{U}_\theta \subset \mathcal{M} \text{ and } \varphi_\theta : \mathcal{U}_\theta \to \mathbb{R}^d,\ \text{diffeomorphism}$$



Riemannian manifolds (1/2)

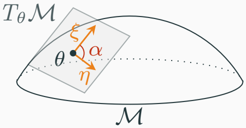
Tangent space $T_\theta\mathcal{M}$ at point $\boldsymbol{\theta}$



- Curve $\gamma : \mathbb{R} \to \mathcal{M}$, $\gamma(0) = \theta$
- Derivative: $\dot{\gamma}(0) = \lim_{t\to 0} \frac{\gamma(t) - \gamma(0)}{t}$
- Tangent space $T_\theta\mathcal{M} = \{\dot{\gamma}(0) : \ \gamma : \mathbb{R} \to \mathcal{M}, \ \gamma(0) = \theta\}$

Equip $T_\theta\mathcal{M}$ with a **Riemannian metric** $\langle\cdot,\cdot\rangle_\theta$ yields a **Riemannian manifold**

- $\langle\cdot,\cdot\rangle_\theta : (T_\theta\mathcal{M} \times T_\theta\mathcal{M}) \to \mathbb{R}$ **inner product** on $T_\theta\mathcal{M}$
  (bilinear, symmetric, positive definite)
- defines length and relative positions of tangent vectors



$$\|\xi\|_\theta^2 = \langle\xi,\xi\rangle_\theta \qquad \alpha(\xi,\eta) = \frac{\langle\xi,\eta\rangle_\theta}{\|\xi\|_\theta \, \|\eta\|_\theta}$$

Riemannian manifolds (2/2)

The Riemannian metric $\langle\cdot,\cdot\rangle_\theta$ induces **a geometry** for $\mathcal{M}$

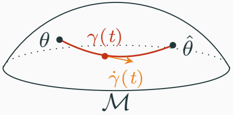
Geodesics $\gamma : [0,1] \to \mathcal{M}$

- generalizes straight lines on $\mathcal{M}$
- curves on $\mathcal{M}$ with zero acceleration: $\frac{D^2\gamma}{dt^2} = 0$

defined by $(\gamma(0), \dot\gamma(0))$ or $(\gamma(0), \gamma(1))$

operator $\frac{D^2}{dt^2}$ depends on $\mathcal{M}$ and $\langle\cdot,\cdot\rangle$.

Riemannian distance $\mathrm{dist}(\theta, \hat\theta) = \int_0^1 \|\dot\gamma(t)\|_{\gamma(t)}\, dt$



distance = length of γ connecting θ and $\hat\theta$

Which metric/geometry to chose ?

The **Fisher information metric** looks like an **ideal driven by the model**

Still, we can chose **alternate metrics** **suited to some needs**

- Availability (**closed-form**) of theoretical objects
- Interesting **invariance** properties
- **Practical results** of the chosen task

Metric	Geodesics	Distance	Retraction	Completeness	Invariance 1	Invariance 2	Perf.
(a)	✗	✗	✓	✓	✗	✓	82%
(b)	✓	✗	✓	✓	✓	✗	**86**%
(c)	✓	✓	✓	✗	✗	✓	79%

Outline

- **Design** ✓
  - Examples of f and $\boldsymbol{\theta}$ from elliptical distributions
  - Remark that $\boldsymbol{\theta} \in \mathcal{M} \implies$ pretext to define Riemannian tools

✗ **Analyze**
  - Intrinsic Cramér-Rao bounds
  - 2 examples of interesting inequalities

- **Solve**
  - Riemannian optimization and geodesic convexity
  - 2 examples where numerical stability is improved

- **Apply**
  - Clustering with Riemannian distances

Cramér-Rao lower bound (CRLB)

CRLB: If $\mathbf{z} \sim f(\mathbf{z}, \boldsymbol{\theta})$, then for $\hat{\boldsymbol{\theta}}$ unbiased estimator of $\boldsymbol{\theta}$ ⚠as a vector

$$\mathbb{E}\left\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right\} \succeq \mathbf{F}^{-1}(\boldsymbol{\theta}) \quad \Rightarrow \quad \text{MSE} \geq \text{Tr}\left\{\mathbf{F}^{-1}(\boldsymbol{\theta})\right\}$$

with the **Fisher information matrix** $\mathbf{F}(\boldsymbol{\theta}) = -\mathbb{E}\left\{ \left. \frac{\partial^2 \ln f(\mathbf{z}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}} \right\}$

Slepian-Bangs formula: if $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Gamma}(\boldsymbol{\theta}))$

$$[\mathbf{F}(\boldsymbol{\theta})]_{ij} = 2\mathfrak{Re}\left\{ \left. \frac{\partial \boldsymbol{\mu}^H(\boldsymbol{\theta})}{\partial \theta_i} \right|_{\boldsymbol{\theta}} \boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}) \left. \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j} \right|_{\boldsymbol{\theta}} \right\} + \text{Tr}\left\{ \boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}) \left. \frac{\partial \boldsymbol{\Gamma}(\boldsymbol{\theta})}{\partial \theta_i} \right|_{\boldsymbol{\theta}} \boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}) \left. \frac{\partial \boldsymbol{\Gamma}(\boldsymbol{\theta})}{\partial \theta_j} \right|_{\boldsymbol{\theta}} \right\}$$

Extension to CES in [Besson13]

"Constrained" CRLB (cCRLB)

Constraints: If elements of $\boldsymbol{\theta}$ are linked by some system

$$h_k(\theta_1, \theta_2, \ldots, \theta_P) = 0,\ k \in [\![1, M]\!] \quad \Leftrightarrow \quad \mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$$

$\mathbf{F}(\boldsymbol{\theta})$ becomes singular $\Rightarrow$ no proper CRLB

cCRLB: we still have [Gorman90]

$$\mathbb{E}\left\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right\} \succeq \mathbf{U}(\boldsymbol{\theta}) \left(\mathbf{U}^T(\boldsymbol{\theta})\mathbf{F}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta})\right)^{-1} \mathbf{U}^T(\boldsymbol{\theta})$$

with $\mathbf{U}(\boldsymbol{\theta})$ such that $\mathbf{H}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$ and $\mathbf{U}^T(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta}) = \mathbf{I}_M$, and $\mathbf{H}(\boldsymbol{\theta}) = \left.\frac{\partial \mathbf{h}(\boldsymbol{\theta})}{\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}}$

But what if $\theta \in \mathcal{M}$?

- **Parameterization and constraints ?**
  - Difficult to have a system of coordinates — e.g. subspaces
  - Difficult (or impossible) to express constraints as $\mathbf{h}(\boldsymbol{\theta})$ — e.g. PSD for $\mathcal{H}_p^{++}$

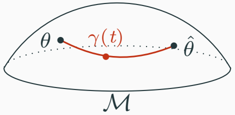
- **Performance measure ?**
  - Can we bound a Riemannian distance rather than the MSE ?
  - Non-trivial function $\Rightarrow$ no Jacobian

$\longrightarrow$ We can turn to the framework of **intrinsic CRLB** (iCRLB)

Riemannian framework of iCRLB

Definitions:

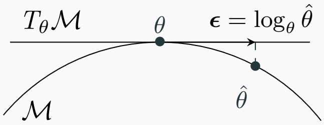
- $\theta \in \mathcal{M}$ with tangent space $T_\theta\mathcal{M}$
- $\hat{\theta} \in \mathcal{M}$ estimate of θ
- $\langle\cdot,\cdot\rangle_\theta$ chosen Riemannian metric
- $\mathrm{dist}(\cdot,\cdot)$ induced Riemannian distance
- $\{\xi_i\}$ corresponding orthonormal basis of $T_\theta\mathcal{M}$



Error measure = $\mathrm{dist}^2(\theta,\hat{\theta})$

Riemannian logarithm $\boldsymbol{\epsilon} = \log_\theta \hat{\theta} \in T_\theta\mathcal{M}$

- Points from θ to $\hat{\theta}$ with $||\log_\theta \hat{\theta}||_\theta^2 = \mathrm{dist}^2(\theta,\hat{\theta})$
- Would be "$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$" in the Euclidean setup
- In coordinates $[\boldsymbol{\epsilon}]_i = \langle \log_\theta \hat{\theta}, \xi_i\rangle_\theta$



Error vector = $\log_\theta \hat{\theta}$

Fisher information metric/matrix

Fisher information metric For $f(\{\mathbf{x}_k\};\theta)$ p.d.f. parameterized by $\boldsymbol{\theta} \in \mathcal{M}$

$$\langle \xi, \xi \rangle_\theta^{\mathrm{FIM}} = -\mathbb{E}\left[\left. \frac{\mathrm{d}^2}{\mathrm{d}t^2} \ln f(\{\mathbf{x}_k\}; \theta + t\xi) \right|_{t=0} \right]$$

Fisher information matrix represented in coordinates $\{\xi_i\}$ by

$$[\mathbf{F}]_{ij} = \langle \xi_i, \xi_j \rangle_\theta^{\mathrm{FIM}}$$

Remarks

- $\langle \cdot, \cdot \rangle_\theta^{\mathrm{FIM}}$ defines a metric for $T_\theta\mathcal{M} \Rightarrow$ **information geometry** for $\mathcal{M}$
- Error measured from $\langle \cdot, \cdot \rangle_\theta$, which can be different

Intrinsic CRLB

Intrinsic CRLB (iRCLB) [Smith05, Boumal14]

Assuming model $f(\{\mathbf{x}_k\};\boldsymbol{\theta})$ and unbiased estimator $\hat{\theta}$, we have

$$\mathbb{E}\left[(\log_\theta \hat{\theta})(\log_\theta \hat{\theta})^H\right] \succeq \mathbf{F}^{-1} - \underbrace{\tfrac{1}{3}\left(\mathbf{F}^{-1}\mathbf{R}_m\left(\mathbf{F}^{-1}\right) + \mathbf{R}_m\left(\mathbf{F}^{-1}\right)\mathbf{F}^{-1}\right) + \mathcal{O}(\lambda_{\max}(\mathbf{F}^{-1})^{2+1/2})}_{\text{Riemannian curvature terms (cf. [Boumal14, Eq.6.6])}}$$

Remarks

- $\mathbf{F}^{-1}$ depends on $\langle\cdot,\cdot\rangle_\theta \Rightarrow$ iCRLB indeed changes w.r.t. d
- Bias terms + more about curvature in [Smith05]
- Neglecting the curvature terms, we have in trace $\mathbb{E}\left\{\text{dist}^2(\hat{\theta},\theta)\right\} \geq \text{Tr}\left\{\mathbf{F}^{-1}\right\}$

"$(\cdot)^{-1}$" inverse of a tensor (defined w.r.t. a metric)

Wrapping up

iCRLB cooking recipe

1. Compute $\langle \xi, \xi \rangle_\theta^{\mathrm{FIM}} = -\mathbb{E}\left[\left. \frac{\mathrm{d}^2}{\mathrm{d}t^2} \ln f(\{\mathbf{x}_k\}; \theta + t\xi) \right|_{t=0} \right]$ and polarization for $\langle \xi_i, \xi_j \rangle_\theta^{\mathrm{FIM}}$
2. Chose the error metric $\langle \cdot, \cdot \rangle_\theta \longrightarrow \begin{cases} \text{error distance } \mathrm{dist} \\ \text{orthonormal basis } \{\xi_i\} \text{ of } T_\theta\mathcal{M} \end{cases}$
3. Compute the Fisher information matrix: $[\mathbf{F}]_{ij} = \langle \xi_i, \xi_j \rangle_\theta^{\mathrm{FIM}}$
4. Bound the expected distance as $\mathbb{E}\left\{\mathrm{dist}^2(\hat{\theta}, \theta)\right\} \geq \mathrm{Tr}\left\{\mathbf{F}^{-1}\right\}$

Interest?

- Bounding other distances: neat formulas, reveals unexpected things (intrinsic bias)
- Parameterization from $T_\theta\mathcal{M} \to$ useful even in the Euclidean case!

Example 1: iCRLB for covariance matrix estimation in CES (1/2)

Model $\mathbf{x} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g)$ with pdf $f(\mathbf{x}) \propto |\mathbf{\Sigma}|^{-1} g\left(\mathbf{x}^H \mathbf{\Sigma}^{-1} \mathbf{x}\right)$, and representation

$$\mathbf{x} \overset{d}{=} \sqrt{\mathcal{Q}} \mathbf{\Sigma}^{1/2} \mathbf{u} \quad \text{with} \begin{cases} \mathbf{u} \text{ uniformly distributed on the unit sphere } \mathbf{u} \sim \mathcal{U}(\mathbb{C}\mathcal{S}^p) \\ \mathcal{Q} \text{ independent modular variate, pdf related to } g \end{cases}$$

Manifold $\mathbf{\Sigma} \in \mathcal{H}_p^{++}$ (Hermitian pd matrices) with tangent space $T_{\mathbf{\Sigma}} \mathcal{H}_p^{++} = \mathcal{H}_p$ (Hermitian matrices)

Error metric: "natural" Riemannian metric and distance for $\mathcal{H}_p^{++}$

$$\langle \xi_i, \xi_j \rangle_{\mathbf{\Sigma}} = \operatorname{Tr}\left\{ \mathbf{\Sigma}^{-1} \xi_i \mathbf{\Sigma}^{-1} \xi_j \right\} \quad \text{inducing} \quad \operatorname{dist}^2_{\mathcal{H}_p^{++}}(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}) = || \log \mathbf{\Sigma}^{-1/2} \hat{\mathbf{\Sigma}} \mathbf{\Sigma}^{-1/2} ||_F^2$$

Example 1: iCRLB for covariance matrix estimation in CES (2/2)

Fisher information metric for CES

aka "affine invariant"

Let $\{\mathbf{x}_i\}_{i=1}^n$ in $\mathbb{C}^p$ with $\mathbf{x} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g)$, then

$$\langle \xi_i, \xi_j \rangle_{\mathbf{\Sigma}}^{\mathrm{FIM}} = n\alpha_g \operatorname{Tr}\left\{\mathbf{\Sigma}^{-1}\xi_i\mathbf{\Sigma}^{-1}\xi_j\right\} + n\beta_g \operatorname{Tr}\left\{\mathbf{\Sigma}^{-1}\xi_i\right\} \operatorname{Tr}\left\{\mathbf{\Sigma}^{-1}\xi_j\right\}$$

with $\alpha_g = 1 - \frac{\mathbb{E}[\mathcal{Q}^2\phi'(\mathcal{Q})]}{M(M+1)}$ and $\beta_g = \alpha - 1$ using $\phi(t) = g'(t)/g(t)$

iCRLB for Σ

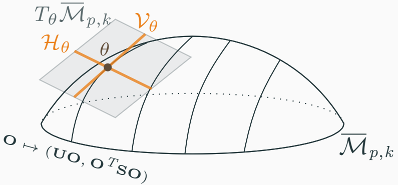
Let $\{\mathbf{z}_i\}_{i=1}^n$ in $\mathbb{C}^p$ with $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g)$

$$\mathbb{E}\left[\mathrm{dist}^2_{\mathcal{H}_p^{++}}\left(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}\right)\right] \geq \frac{1}{n}\left(\frac{p^2-1}{\alpha_g} + \frac{1}{\alpha_g(p+1)-p}\right)$$

Example 2: spiked model (PCA) in CES

Model: $\mathbf{x} \sim \mathcal{CES}(\mathbf{0}, \mathbf{H} + \mathbf{I}, g)$, with $\mathbf{H} \in \mathcal{H}_{p,k}^{+}$
(H-psd of rank k)

Manifold: $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H \in \mathcal{H}_{p,k}^{+}$ as $(\mathrm{St}(p,k) \times \mathcal{H}_k^{++})/\mathcal{U}_k$



Error metric:

$$\langle \bar{\xi}, \bar{\eta} \rangle_{\bar{\theta}} = \underbrace{\Re\mathfrak{e}(\mathrm{Tr}(\xi_{\mathbf{U}}^H(\mathbf{I}_p - \tfrac{1}{2}\mathbf{U}\mathbf{U}^H)\boldsymbol{\eta}_{\mathbf{U}}))}_{\text{canonical on } \mathrm{St}(p,k)} + \underbrace{\alpha\mathrm{Tr}(\mathbf{\Sigma}^{-1}\boldsymbol{\xi}_{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\boldsymbol{\eta}_{\mathbf{\Sigma}}) + \beta\mathrm{Tr}(\mathbf{\Sigma}^{-1}\boldsymbol{\xi}_{\mathbf{\Sigma}})\mathrm{Tr}(\mathbf{\Sigma}^{-1}\boldsymbol{\eta}_{\mathbf{\Sigma}})}_{\text{affine invariant on } \mathcal{H}_k^{++}}$$

iCRLB for subspace

Let $\{\mathbf{z}_i\}_{i=1}^n$ in $\mathbb{C}^p$ with $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{U}\mathrm{diag}(\{\sigma_r\}_{r=1}^k)\mathbf{U}^H + \mathbf{I}, g)$

$$\mathbb{E}\left[\mathrm{dist}^2_{\mathcal{G}_{p,k}}\left(\mathrm{span}(\hat{\mathbf{U}}), \mathrm{span}(\mathbf{U})\right)\right] \geq \frac{p-k}{n\alpha_g} \sum_{r=1}^{k} \frac{1+\sigma_r}{\sigma_r^2}$$

Outline

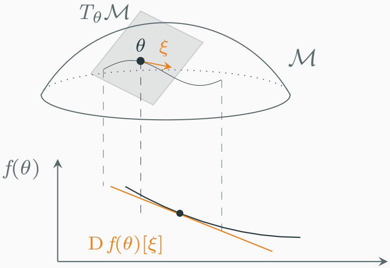
- **Design** ✓
  - Examples of f and $\boldsymbol{\theta}$ from elliptical distributions
  - Remark that $\boldsymbol{\theta} \in \mathcal{M} \implies$ pretext to define Riemannian tools
- **Analyze** ✓
  - Intrinsic Cramér-Rao bounds
  - 2 examples of interesting inequalities

✗ **Solve**
  - Riemannian optimization and geodesic convexity
  - 2 examples where numerical stability is improved
- **Apply**
  - Clustering with Riemannian distances

Riemannian optimization

$$\underset{\theta \in \mathcal{M}}{\text{minimize}} \quad f(\theta)$$

Riemannian optimization: a framework for optimization on $\mathcal{M}$ equipped with $\langle \cdot, \cdot \rangle$.



Descent direction of f at θ:

$$\xi \in T_\theta \mathcal{M}, \quad \mathrm{D}\, f(\theta)[\xi] < 0$$

Riemannian gradient of f at θ:

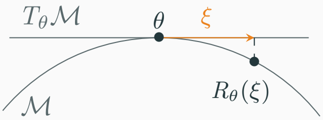
$$\langle \operatorname{grad} f(\theta), \xi \rangle_\theta = \mathrm{D}\, f(\theta)[\xi]$$

Riemannian optimization

[?]

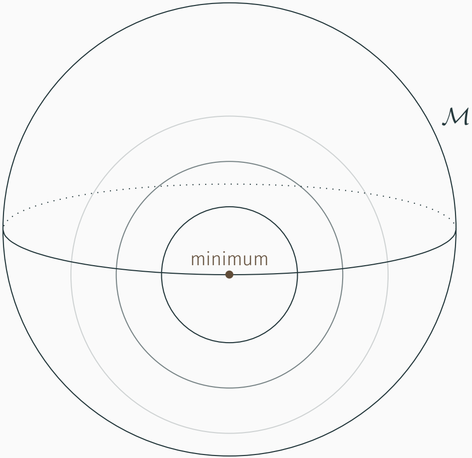
Main ingredients

- Descent direction: $\xi \in T_\theta\mathcal{M}$ so that $\langle \operatorname{grad} f(\theta), \xi \rangle_\theta < 0$
- **Retraction** of ξ on $\mathcal{M}$ (smooth mapping)

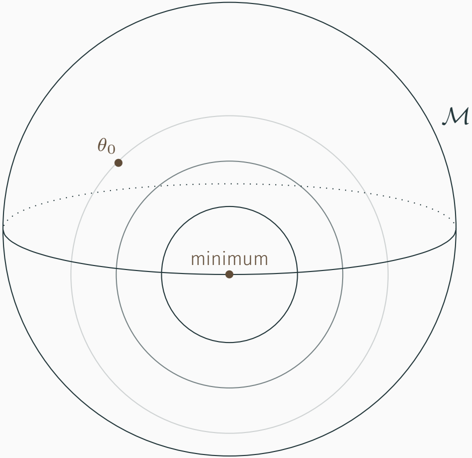


Flexibility: metric, retraction, descent method (gradient, conjugate gradient, BFGS...)

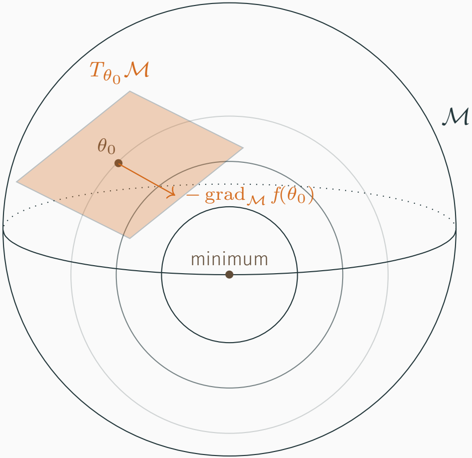
Example: Riemannian gradient descent $\theta_{i+1} = R_{\theta_i}(-t_i \operatorname{grad} f(\theta_i))$



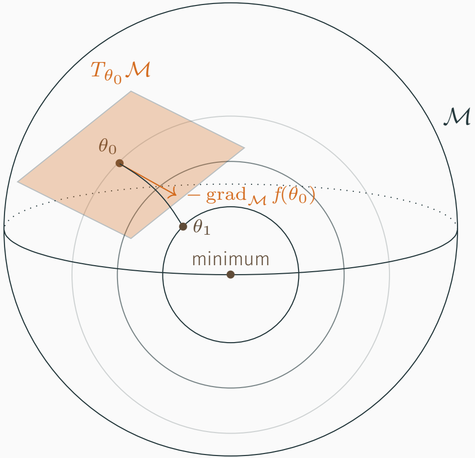
Example: Riemannian gradient descent $\theta_{i+1} = R_{\theta_i}(-t_i \,\mathrm{grad}\, f(\theta_i))$



Example: Riemannian gradient descent $\theta_{i+1} = R_{\theta_i}(-t_i \operatorname{grad} f(\theta_i))$



Example: Riemannian gradient descent $\theta_{i+1} = R_{\theta_i}(-t_i \operatorname{grad} f(\theta_i))$

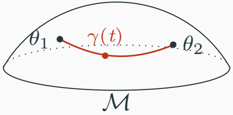


Geodesic convexity (g-convexity)

f is **g-convex** if $\forall\ \theta_1, \theta_2 \in \mathcal{M}$, f is convex on geodesic $\gamma(t)$, i.e

$$f(\gamma(t)) \leq t\, f(\theta_1) + (1-t)\, f(\theta_2)$$

If so, then any local minimizer is a global minimizer.



Example: CES log-likelihoods

$$\mathcal{L}(\boldsymbol{\Sigma}) = \sum_{i=1}^{n} \ln g\left(\mathbf{x}_i^H \boldsymbol{\Sigma}^{-1} \mathbf{x}_i\right) + n \ln |\boldsymbol{\Sigma}|$$

are g-convex following the geodesics $\boldsymbol{\Sigma}(t) = \boldsymbol{\Sigma}_1^{1/2} \left(\boldsymbol{\Sigma}_1^{-1/2} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1/2}\right)^t \boldsymbol{\Sigma}_1^{1/2}$

has been useful to prove uniqueness of (regularized) M-estimators

Example 1: robust mean and covariance estimation (1/3)

Jointly estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for $\mathbf{x} \sim \mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

***M*-estimators** of location and scatter

$$\boldsymbol{\mu} = \left(\sum_{i=1}^{n} u_1(t_i)\right)^{-1} \sum_{i=1}^{n} u_1(t_i)\mathbf{x}_i \qquad \boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} u_2(t_i)(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^H$$

where $t_i \triangleq (\mathbf{x}_i - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})$, and u_1, u_2 respect conditions in [Maronna76]

Tyler's estimator

$$\boldsymbol{\mu} = \left(\sum_{i=1}^{n} \frac{1}{\sqrt{t_i}}\right)^{-1} \sum_{i=1}^{n} \frac{\mathbf{x}_i}{\sqrt{t_i}} \qquad \boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \frac{(\mathbf{x}_i - \boldsymbol{\mu})(-\boldsymbol{\mu})^H}{t_i}$$

Possible fixed-point issues when $t_i \simeq 0$

Example 1: robust mean and covariance estimation (2/3)

Alternatively when $\boldsymbol{\mu} = \mathbf{0}$: Tyler's estimator ⇔ **MLE for scaled Gaussian** $\mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \tau_i \boldsymbol{\Sigma})$

Transposed to non-zero mean $\mathbf{x}_i \sim \mathcal{CN}(\boldsymbol{\mu}, \tau_i \boldsymbol{\Sigma})$

$$\underset{\boldsymbol{\mu}, \{\tau_i\}_{i=1}^n, \boldsymbol{\Sigma}}{\text{maximize}} \quad \sum_{i=1}^{n} \left[\ln |\tau_i \boldsymbol{\Sigma}| + \frac{(\mathbf{x}_i - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})}{\tau_i} \right]$$

yields

$$\boldsymbol{\mu} = \left(\sum_{i=1}^{n} \frac{1}{t_i} \right)^{-1} \sum_{i=1}^{n} \frac{\mathbf{x}_i}{t_i} \qquad \boldsymbol{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \frac{(\mathbf{x}_i - \boldsymbol{\mu})(-\boldsymbol{\mu})^H}{t_i}$$

slightly different but fixed-point iterations diverge in practice!

Example 1: robust mean and covariance estimation (3/3)

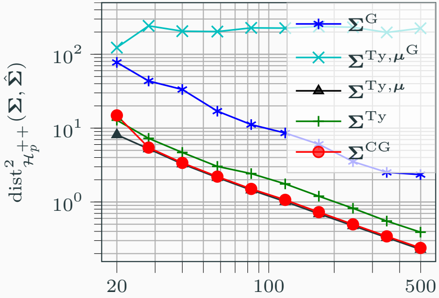
Product manifold $\mathcal{M}_{p,n} \in \mathbb{C}^p \times (\mathbb{R}_\star^+)^n \times \mathcal{SH}_p^{++}$ with ***decoupled* metric**

$(\mathcal{H}_p^{++} \cap \det = 1)$

$$\langle \xi, \eta \rangle_\theta^{\mathcal{M}_{p,n}} = \underbrace{\mathfrak{Re}\{\boldsymbol{\xi}_\mu^H \boldsymbol{\eta}_\mu\}}_{\text{canonical on } \mathbb{C}^p} + \underbrace{(\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\xi}_{\boldsymbol{\tau}})^T (\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\eta}_{\boldsymbol{\tau}})}_{\text{canonical on } (\mathbb{R}_\star^+)^n} + \underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}_{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\eta}_{\boldsymbol{\Sigma}})}_{\text{Natural Riem. on } \mathcal{SH}_p^{++}}$$

And resulting:

- Riemannian gradient descent
- Surprisingly stable and accurate estimator
- Still... slow convergence
- Faster with information geometry to appear!



Example 2: robust estimator for spiked models in CES (1/2)

Spiked Tyler's estimator

$$\begin{array}{ll} \underset{\boldsymbol{\Sigma}}{\text{minimize}} & \frac{p}{n}\sum_{i=1}^{n} \ln\left(\mathbf{x}_i^H \boldsymbol{\Sigma}^{-1}\mathbf{x}_i\right) + \ln|\boldsymbol{\Sigma}| \\ \text{subject to} & \boldsymbol{\Sigma} = \mathbf{H} + \sigma^2\mathbf{I}, \text{ with } \mathbf{H} \in \mathcal{H}_{p,k}^{+} \end{array}$$

Existing **MM algorithm** [Sun16]

1. Usual fixed point iteration

$$\boldsymbol{\Sigma}_{t+1/2} = \frac{p}{n}\sum_{i=1}^{n} \frac{\mathbf{x}_i\mathbf{x}_i^H}{\mathbf{x}_i^H\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_i}$$

2. Projection on the structured set

$$\boldsymbol{\Sigma}_{t+1} = \mathcal{P}_{\mathcal{H}_{p,k}^{+}}\left(\boldsymbol{\Sigma}_{t+1/2}\right)$$

where $\mathcal{P}_{\mathcal{H}_{p,k}^{+}}$ averages the last $p-k$ eigenvalues (SVD)

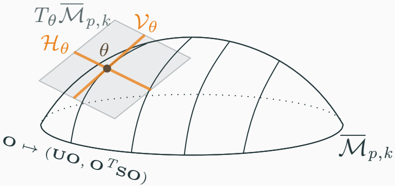
can diverge with small n

Example 2: robust estimator for spiked models in CES (2/2)

Riemannian optimization for

$$\underset{\mathbf{H}\in\mathcal{H}_{p,k}^{+}}{\text{minimize}} \quad \mathcal{L}_{\text{Ty}}(\mathbf{H}+\mathbf{I})$$

with $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H \in (\text{St}(p,k) \times \mathcal{H}_k^{++})/\mathcal{U}_k$

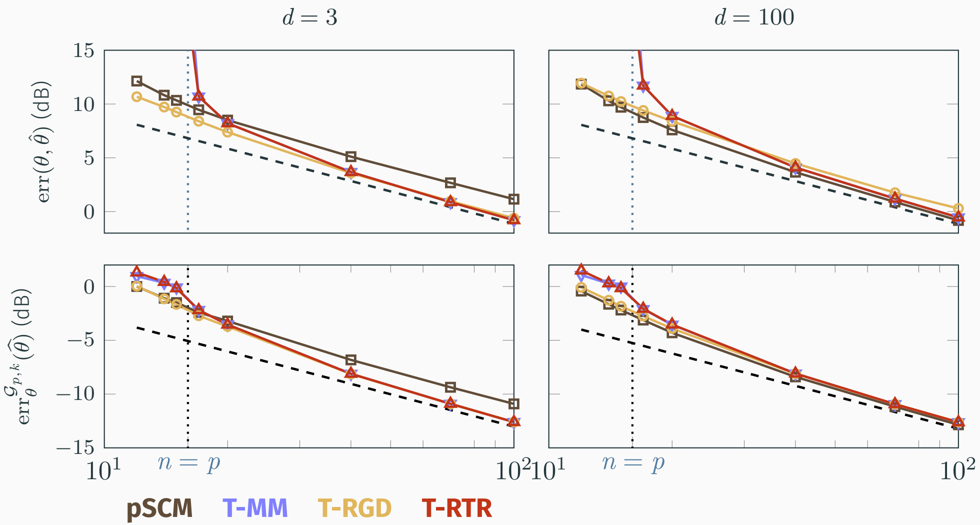


using the metric

$$\langle \bar{\xi}, \bar{\eta} \rangle_{\bar{\theta}} = \underbrace{\mathfrak{Re}(\text{Tr}(\xi_{\mathbf{U}}^H(\mathbf{I}_p - \tfrac{1}{2}\mathbf{U}\mathbf{U}^H)\boldsymbol{\eta}_{\mathbf{U}}))}_{\text{canonical on St}(p,k)} + \underbrace{\alpha\text{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}_{\boldsymbol{\Sigma}}) + \beta\text{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_{\boldsymbol{\Sigma}})\text{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}_{\boldsymbol{\Sigma}})}_{\text{affine invariant on } \mathcal{H}_k^{++}}$$

$\longrightarrow$ Riemannian **gradient descent (T-RGD)** and **trust region (T-RTR)** algorithms

Numerical illustrations: t-distribution $p = 16$, $k = 8$, SNR$\simeq 15$dB



Outline

- **Design** ✓
  - Examples of f and $\boldsymbol{\theta}$ from elliptical distributions
  - Remark that $\boldsymbol{\theta} \in \mathcal{M} \implies$ pretext to define Riemannian tools
- **Analyze** ✓
  - Intrinsic Cramér-Rao bounds
  - 2 examples of interesting inequalities
- **Solve** ✓
  - Riemannian optimization and geodesic convexity
  - 2 examples where numerical stability is improved

✗ **Apply**

  - Clustering with Riemannian distances

Clustering problem

Mixture model: observations follow $(\mathbf{x}|\text{class } k) \sim f(x, \theta_k)$ with K possible classes

Clustering: from **unlabeled** data $\{\mathbf{x}_i\}_{i=1}^n$ find the partition $\{\{\mathbf{x}_i^k\}_{i=1}^{n_k}\}_{k=1}^K$

Issues:

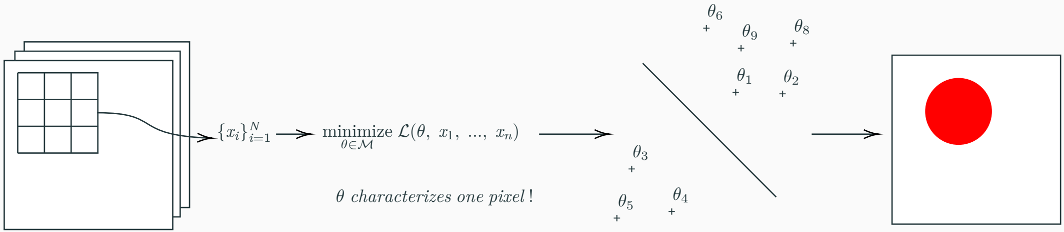
- Statistical ideal would be the **EM algorithm** $\to$ no time for that!
- More accurate model could involve f_k $\to$ need for robustness to mismatches
- Elements in θ_k might be non-discriminating

A standard solution is to cluster intermediate **features**

(aka descriptors)

Feature clustering pipeline with a geometric twist

Step 1: **sliding window** Step 2: **parameter estimation** Step 3: **parameter clustering**



Riemannian approach for $\theta \in \mathcal{M}$: **transpose** clusterings algorithm using

- **Information geometry** of model $\mathbf{x} \sim f(x, \theta)$
- **Distances** $\text{dist}^2(\theta_i, \theta_j)$ and **Riemannian means** $\underset{\theta}{\text{argmin}} \sum_{i=1}^{j} \text{dist}^2(\theta, \theta_i)$

computed with Riem. opt.

An example on Indian pines data set

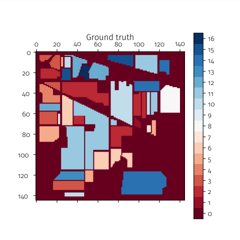
Plain K-means++, compared to Riemmanian counterparts from two models:

Centered Gaussian (GMM)

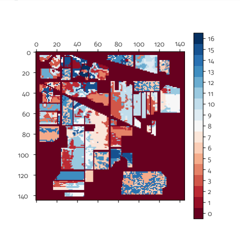
- $\mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$
- θ = covariance matrix $\mathbf{\Sigma}$
- Natural distance $\mathrm{dist}_{\mathcal{H}_p^{++}}$

Probabilistic PCA with SG signals

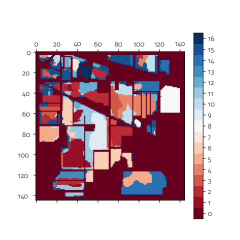
- $\mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \tau_i \mathbf{U}\mathbf{U}^{\mathbf{H}} + \mathbf{I})$
- θ = subspace $\mathrm{span}(\mathbf{U})$ + textures $\{\tau_i\}_{i=1}^n$
- Decoupled distance on $\mathrm{Gr}_p^k \times \mathbb{R}^n$



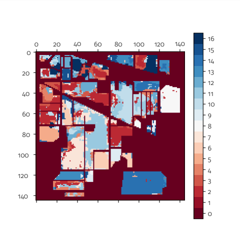
Ground truth



$\mathbb{R}^p$: OA $= 31.2\%$



$\mathcal{H}_p^{++}$: OA $= 45.2\%$



$\mathrm{Gr}_p^k \times \mathbb{R}^n$: OA $= 47.2\%$

Outline

- **Design** ✓
  - Examples of f and $\boldsymbol{\theta}$ from elliptical distributions
  - Remark that $\boldsymbol{\theta} \in \mathcal{M} \implies$ pretext to define Riemannian tools
- **Analyze** ✓
  - Intrinsic Cramér-Rao bounds
  - 2 examples of interesting inequalities
- **Solve** ✓
  - Riemannian optimization and geodesic convexity
  - 2 examples where numerical stability is improved
- **Apply** ✓
  - Clustering with Riemannian distances

Perspectives in regularization

- **Intrinsic bias** $\mathrm{b}(\hat{\theta}) = \mathbb{E}\left[\log_{\theta} \hat{\theta}\right] \rightarrow$ **counter-intuitive bias-variance paradigm**

$$\mathrm{b}(\boldsymbol{\Sigma}_{\mathrm{SCM}}) = \mathbb{E}\left[\log_{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_{\mathrm{SCM}}\right] = \boldsymbol{\Sigma}^{1/2}\mathbb{E}\left[\log(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}_{\mathrm{SCM}}\boldsymbol{\Sigma}^{-1/2})\right]\boldsymbol{\Sigma}^{1/2} = \textbf{not zero!}$$

- **Geodesic shrinkage** or not?

$$\boldsymbol{\Sigma}_g(t) = \mathbf{T}_1^{1/2}\left(\mathbf{T}_1^{-1/2}\hat{\boldsymbol{\Sigma}}\mathbf{T}_1^{-1/2}\right)^t\mathbf{T}_1^{1/2} \qquad \text{versus} \qquad \boldsymbol{\Sigma}_L(t) = t\hat{\boldsymbol{\Sigma}} + (1-t)\mathbf{T}$$

- Can we do **Stein's type** regularization (shrink eigenvalues) for $\mathbb{E}\left[\mathrm{dist}_{\mathcal{H}_p^{++}}(\tilde{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})\right]$?

References

On CRLBs with constraints

"*Lower bounds for parametric estimation with constraints*," IT, 1990
John D. Gorman, Alfred O. Hero,

"*A simple derivation of the constrained multiple parameter Cramér-Rao bound*," TSP, 1993
Thomas L. Marzetta

"*On the Cramér-Rao bound under parametric constraints*," SPL, 1998
Petre Stoica, Boon Chong Ng

On iCRLBs

"*Covariance, subspace, and intrinsic Cramér-Rao bounds*," TSP, 2005
Steven T. Smith

"*On intrinsic Cramér-Rao bounds for Riemannian submanifolds and quotient manifolds*," TSP, 2013
Nicolas Boumal

"*Optimization and estimation on manifolds*," Doctoral dissertation, Catholic University of Louvain
Nicolas Boumal

References

On Riemannian optimization

"*Optimization algorithms on matrix manifolds*," Princeton University Press, 2009
Pierre-Antoine Absil, Robert Mahony, Rodolphe Sepulchre

"*An introduction to optimization on smooth manifolds*," available online, 2020
Nicolas Boumal

On Riemannian geometry for low-rank PSD matrices

"*Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank*," SIMAX, 2009
Silvere Bonnabel and Rodolphe Sepulchre.

"*Regression on fixed-rank positive semidefinite matrices: a Riemannian approach*," JMLR, 2011
Gilles Meyer, Silvere Bonnabel, Rodolphe Sepulchre

"*A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank*," IMAJNA, 2012
Bart Vandereycken, Pierre-Antoine Absil, Stefan Vandewalle

"*Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices. Technical*," SIMAX, 2019
Estelle Massart, Pierre-Antoine Absil

References

On CES

"*Complex elliptically symmetric distributions: Survey, new results and applications*," TSP, 2012
Esa Ollila, David E. Tyler, Visa Koivunen, H. Vincent Poor

"*Geodesic convexity and covariance estimation*," TSP, 2012
"*Unified framework to regularized covariance estimation in scaled gaussian models*," TSP, 2012
Ami Wiesel

"*Regularized M-estimators of scatter matrix*," TSP, 2014
Esa Ollila, David E. Tyler

Work presented in this talk

"*Intrinsic Cramér–Rao bounds for scatter and shape matrices estimation in CES distributions*," SPL, 2018.
Arnaud Breloy, Guillaume Ginolhac, Alexandre Renaux, Florent Bouchard

"*A Riemannian Framework for Low-Rank Structured Elliptical Models*," TSP, 2021.
Florent Bouchard, Arnaud Breloy, Guillaume Ginolhac, Alexandre Renaux, Frederic Pascal

"*A Tyler-Type Estimator of Location and Scatter Leveraging Riemannian Optimization*," ICASSP 2021.
Antoine Collas, Florent Bouchard, Arnaud Breloy, Guillaume Ginolhac, Chengfang Ren, Jean-Philippe Ovarlez

"*Probabilistic PCA from Heteroscedastic Signals: Geometric Framework and Application to Clustering*"
Antoine Collas, Florent Bouchard, Arnaud Breloy, Guillaume Ginolhac, Chengfang Ren, Jean-Philippe Ovarlez