Riemannian geometry for data analysis:

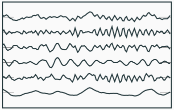
application to blind source separation and low-rank structured covariance matrices

Florent Bouchard (L2S – CNRS, Univ. Paris-Saclay, CentraleSupélec)

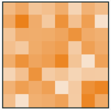
Séminaire Palaisien – 3 Mars 2021

Introduction

- Characterize data $\boldsymbol{x}$ with some parameters $\boldsymbol{\theta}$



$\boldsymbol{x}$



$\boldsymbol{\theta}$

- To estimate $\boldsymbol{\theta}$, optimization problem:

$$\underset{\boldsymbol{\theta}\in\mathcal{M}}{\operatorname{argmin}} \quad f(\boldsymbol{x},\boldsymbol{\theta})$$

  - $f : \mathcal{X} \times \mathcal{M} \to \mathbb{R}$, cost function corresponding to the model
  - $\boldsymbol{x}$ and $\boldsymbol{\theta}$ might possess a structure $\Rightarrow$ $\mathcal{X}$ and $\mathcal{M}$ are manifolds

Introduction – Riemannian geometry

Geometry: metric, curves, distance or divergence

treat difficult problems, complicated to handle with other approaches

- modeling: design appropriate cost functions for the considered model

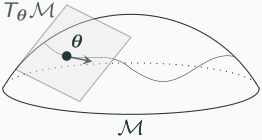
  exploit the structure of data $\boldsymbol{x}$, *e.g.* centers of mass

- optimization: generic convex and non-convex methods, modularity

  naturally takes into account structure of parameters $\boldsymbol{\theta}$

- performance analysis: error measures and associated performance bounds

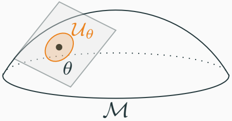
  captures the geometrical structure of the problem



Riemannian geometry and optimization

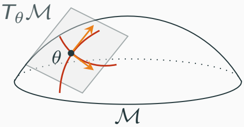
- Manifold $\mathcal{M}$: locally diffeomorphic to $\mathbb{R}^d$, with $\dim(\mathcal{M}) = d$, *i.e.*

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



- 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 $T_\theta\mathcal{M} = \{\dot\gamma(0) :\ \gamma : \mathbb{R} \to \mathcal{M},\ \gamma(0) = \theta\}$



Riemannian geometry

[Absil et al., 2008]

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

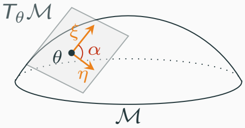
$$\|\xi\|_\theta^2 = \langle \xi, \xi \rangle_\theta \qquad \alpha(\xi, \eta) = \frac{\langle \xi, \eta \rangle_\theta}{\|\xi\|_\theta \|\eta\|_\theta}$$

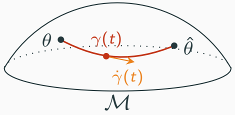
- Geodesics $\gamma : [0,1] \to \mathcal{M}$
  - generalizes straight lines to manifolds – defined by $(\gamma(0), \dot\gamma(0))$ or $(\gamma(0), \gamma(1))$
  - curves on $\mathcal{M}$ with zero acceleration: $\frac{D^2\gamma}{dt^2} = 0$

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

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

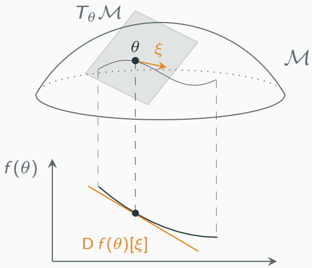
γ: geodesic connecting θ and $\hat\theta$ $\Rightarrow$ distance = length of γ





$$\underset{\theta \in \mathcal{M}}{\operatorname{argmin}} \quad f(\theta)$$

- framework for optimization on manifold $\mathcal{M}$ equipped with metric $\langle \cdot, \cdot \rangle$.



- descent direction of f at θ:

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

- gradient of f at θ:

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

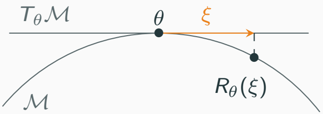
Riemannian optimization

[Absil et al., 2008]

- minimize f on $\mathcal{M}$ from θ:
  - descent direction $\xi \in T_\theta\mathcal{M}$

$$\langle \mathrm{grad}\, f(\theta), \xi \rangle_\theta < 0$$

  - retraction of ξ on $\mathcal{M}$



  - reiterate until critical point: $\mathrm{grad}\, f(\theta) = 0$

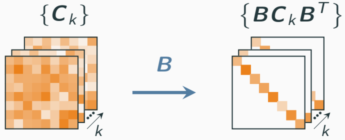
- example of optimization algorithm: gradient descent

$$\theta_{i+1} = R_{\theta_i}(-t_i\, \mathrm{grad}\, f(\theta_i))$$

Approximate joint diagonalization for blind source separation

Problem: approximate joint diagonalization

- *data*: set $\{\boldsymbol{C}_k\}$ of $n \times n$ symmetric positive definite matrices
- *goal*: find joint diagonalizer $\boldsymbol{B}$ of matrices $\boldsymbol{C}_k$



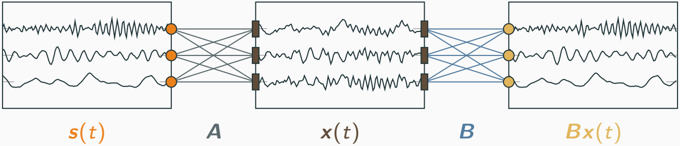
- *formulation*: non-singular matrix $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ such that:

$$\underset{\boldsymbol{B}}{\operatorname{argmin}} \quad f(\boldsymbol{B}, \{\boldsymbol{C}_k\})$$

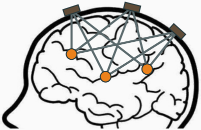
f – diagonality criterion

Application: blind source separation

- *instantaneous linear mixing*: $\boldsymbol{x}(t) = \boldsymbol{A}\,\boldsymbol{s}(t)$ [Comon and Jutten, 2010]
- *goal*: retrieve $\boldsymbol{A}$ and $\boldsymbol{s}(t)$ knowing $\boldsymbol{x}(t)$

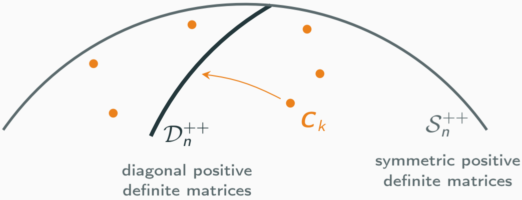


- used in many engineering fields such as electroencephalography



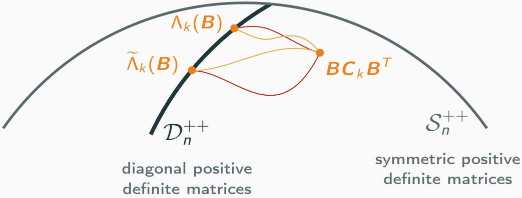
Geometrical modeling of the problem

- *goal*: get $\boldsymbol{C}_k$ as close as possible to $\mathcal{D}_n^{++}$ in $\mathcal{S}_n^{++}$
- *diagonality measure of $\boldsymbol{B}\boldsymbol{C}_k\boldsymbol{B}^T$*: relative position to $\mathcal{D}_n^{++}$



Proposed geometrical model

- *appropriate criterion*: $f(\boldsymbol{B}, \{\boldsymbol{C}_k\})) = \sum_k d(\boldsymbol{B}\boldsymbol{C}_k\boldsymbol{B}^T, \Lambda_k(\boldsymbol{B}))$
  - $\Lambda_k(\boldsymbol{B})$: target diagonal matrices
  - $d(\cdot,\cdot)$: divergence on $\mathcal{S}_n^{++}$



- natural choice for $\Lambda_k(\boldsymbol{B})$:

$$\Lambda_k(\boldsymbol{B}) = \underset{\Lambda \in \mathcal{D}_n^{++}}{\operatorname{argmin}} \; d(\boldsymbol{B}\boldsymbol{C}_k\boldsymbol{B}^T, \Lambda)$$

Considered divergences

- least squares criterion: Euclidean distance on $\mathcal{S}_n^{++}$

[Cardoso and Souloumiac, 1993]

$$\delta_{\mathbf{F}}^2(\boldsymbol{C}, \Lambda) = \|\boldsymbol{C} - \Lambda\|_{\mathbf{F}}^2 \qquad \Lambda = \mathrm{ddiag}(\boldsymbol{C})$$

- log-likelihood cirterion: left Kullback-Leibler divergence [Pham, 2000]

$$d_{\ell\mathbf{KL}}(\boldsymbol{C}, \Lambda) = d_{\mathbf{KL}}(\boldsymbol{C}, \Lambda) \qquad \Lambda = \mathrm{ddiag}(\boldsymbol{C})$$

where $\quad d_{\mathbf{KL}}(\boldsymbol{P}, \boldsymbol{S}) = \mathrm{trace}(\boldsymbol{P}\boldsymbol{S}^{-1} - \boldsymbol{I}_n) - \log\det(\boldsymbol{P}\boldsymbol{S}^{-1})$

- other measures obtained from $d_{\mathbf{KL}}(\cdot, \cdot)$:
  - right measure:

$$d_{r\mathbf{KL}}(\boldsymbol{C}, \Lambda) = d_{\mathbf{KL}}(\Lambda, \boldsymbol{C}) \qquad \Lambda = \mathrm{ddiag}(\boldsymbol{C}^{-1})^{-1}$$

  - symmetrized measure:

$$d_{s\mathbf{KL}}(\boldsymbol{C}, \Lambda) = \frac{1}{2}(d_{\mathbf{KL}}(\boldsymbol{C}, \Lambda) + d_{\mathbf{KL}}(\Lambda, \boldsymbol{C})) \qquad \Lambda = \mathrm{ddiag}(\boldsymbol{C})^{1/2}\,\mathrm{ddiag}(\boldsymbol{C}^{-1})^{-1/2}$$

Considered divergences

- natural Riemannian distance: [Skovgaard, 1984], [Bhatia, 2009]

$$\delta^2_{\mathbf{R}}(\boldsymbol{C}, \Lambda) = \left\| \log(\Lambda^{-1/2} \boldsymbol{C} \Lambda^{-1/2}) \right\|^2_{\mathbf{F}} \qquad \operatorname{ddiag}(\log(\boldsymbol{C}^{-1}\Lambda)) = 0_n$$

- log-Euclidean distance: [Arsigny et al., 2007]

$$\delta^2_{\mathbf{LE}}(\boldsymbol{C}, \Lambda) = \| \log(\boldsymbol{C}) - \log(\Lambda) \|^2_{\mathbf{F}} \qquad \Lambda = \exp(\operatorname{ddiag}(\log(\boldsymbol{C})))$$

- Bhattacharyya distance: [Chebbi and Moakher, 2012], [Sra, 2013]

$$\delta^2_{\mathbf{B}}(\boldsymbol{C}, \Lambda) = 4 \log \frac{\det((\boldsymbol{C} + \Lambda)/2)}{\det(\boldsymbol{C})^{1/2} \det(\Lambda)^{1/2}} \qquad 2 \operatorname{ddiag}((\boldsymbol{C} + \Lambda)^{-1}) = \Lambda^{-1}$$

- Wasserstein distance: [Villani, 2008], [Bhatia et al., 2017]

$$\delta^2_{\mathbf{W}}(\boldsymbol{C}, \Lambda) = \operatorname{trace}\left(\frac{1}{2}(\boldsymbol{C} + \Lambda) - (\Lambda^{1/2} \boldsymbol{C} \Lambda^{1/2})^{1/2} \right) \qquad \operatorname{ddiag}((\Lambda^{1/2} \boldsymbol{C} \Lambda^{1/2})^{1/2}) = \Lambda$$

First experiment: simulated data

$$\boldsymbol{C}_k = \boldsymbol{A}\Lambda_k\boldsymbol{A}^T + \frac{1}{\sigma}\boldsymbol{E}_k\Delta_k\boldsymbol{E}_k^T$$

- $\boldsymbol{A}$, $\boldsymbol{E}_k$: random elements, standard normal distribution

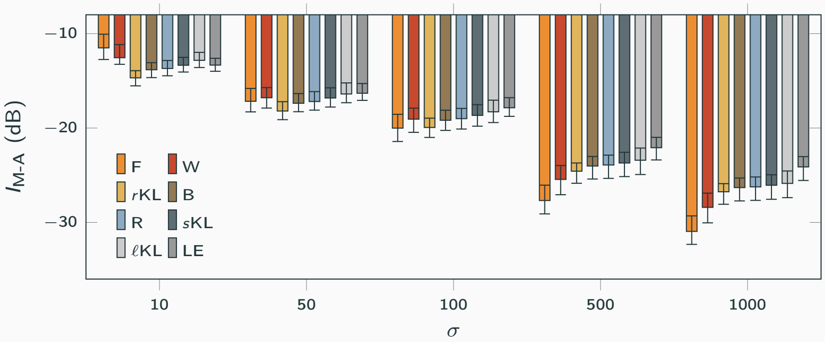
  conditioning of $\boldsymbol{A}$ with respect to inversion in $[1, 10]$
- Λ_k, Δ_k: diagonal – p^{th} element, χ^2 distribution with expectancy 1 and divided by p
- σ : signal to noise ratio

• performance measure: Moreau-Amari criterion [Moreau and Macchi, 1994]

measure degree of similarity of $\boldsymbol{BA}$ with matrix of the form $\boldsymbol{P}\Sigma$

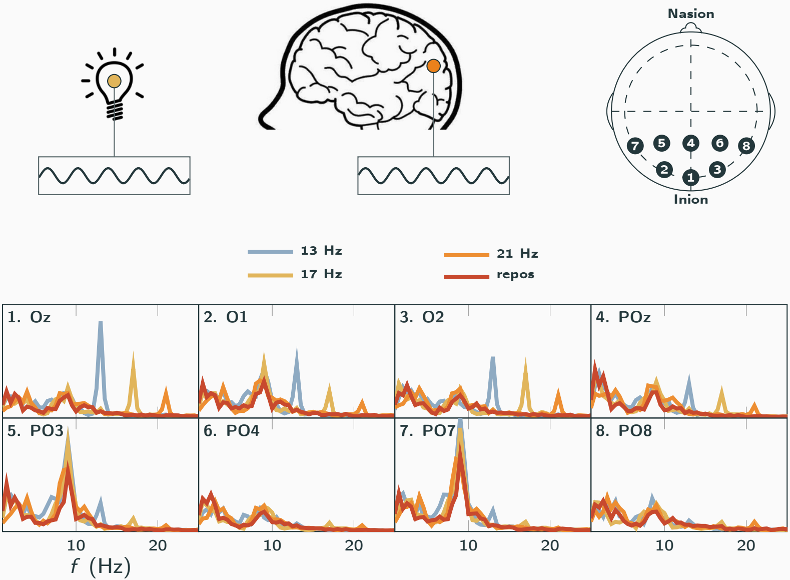
$\boldsymbol{P}$ permutation matrix – Σ non-singular diagonal matrix

First experiment: comparisons of divergences

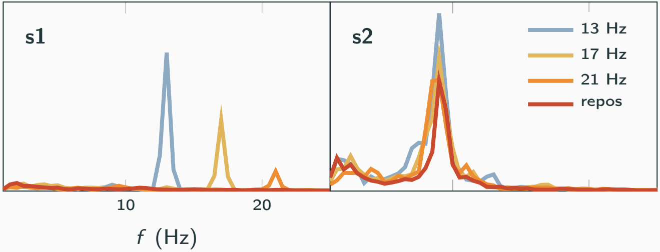


medians, first and ninth deciles (error bars) estimated on 50 trials

Second experiment: electroencephalographic data



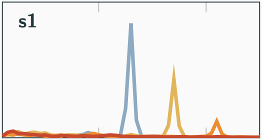
Obtained sources with proposed methods



performance measure $I_{\mathrm{SSVEP}}(f) \in [0, 1]$ for class f

ratio between power at f and total power

$I_{\mathrm{SSVEP}}(f)$	F	W	rKL	B	R	sKL	ℓKL	LE
13 Hz	0, 95	0, 95	0, 96	0, 96	0, 96	0, 96	0, 96	0, 96
17 Hz	0, 87	0, 89	0, 91	0, 91	0, 91	0, 91	0, 91	0, 91
21 Hz	0, 50	0, 54	0, 60	0, 60	0, 60	0, 60	0, 60	0, 60



Intrinsic Cramér-Rao bound for low-rank structured elliptical models

Intrinsic Cramér-Rao bound

- $\boldsymbol{x} \sim \mathcal{L}(\theta)$, $\theta \in \mathcal{M}$, with log-likelihood $L_{\boldsymbol{x}} : \mathcal{M} \to \mathbb{R}$
- given $\boldsymbol{x}$, estimation problem:

$$\hat{\theta} = \operatorname*{argmin}_{\theta \in \mathcal{M}} \quad -L_{\boldsymbol{x}}(\theta)$$

- Cramèr-Rao bound of an unbiased estimator $\hat{\theta}$ of θ :

$$\mathbb{E}_{\hat{\theta}}[\mathrm{Err}(\theta, \hat{\theta})] \succeq \boldsymbol{F}^{-1} \quad \Rightarrow \quad \mathrm{MSE} = \mathbb{E}_{\hat{\theta}}[\mathrm{err}(\theta, \hat{\theta})] \geq \mathrm{trace}(\boldsymbol{F}^{-1})$$

$\boldsymbol{F}$: Fisher information matrix

Intrinsic Cramér-Rao bound

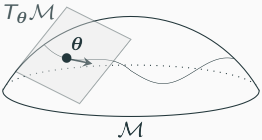
$$\mathbb{E}_{\hat{\theta}}[\mathrm{err}(\theta, \hat{\theta})] \geq \mathrm{trace}(\boldsymbol{F}^{-1})$$

- classical case: $\mathcal{M} = \mathbb{R}^d$

$$\mathrm{err}(\theta, \hat{\theta}) = \|\theta - \hat{\theta}\|_2^2 \qquad\qquad F_{ij} = -\mathbb{E}_{\boldsymbol{x}}\left[\frac{\partial^2 L_{\boldsymbol{x}}(\theta)}{\partial\theta_i \partial\theta_j}\right]$$

$\Rightarrow$ constraints, invariances of the model not taken into account

- generalization to Riemannian manifold $\mathcal{M}$ [Smith, 2005, Boumal, 2013]
  - $\mathrm{err}(\theta, \hat{\theta}) = \delta^2_{\mathcal{M}}(\theta, \hat{\theta})$
    $\delta_{\mathcal{M}}$ distance on $\mathcal{M}$ associated to $\langle\cdot,\cdot\rangle^{\mathcal{M}}_{\cdot}$
  - $F_{ij} = \langle e_i, e_j \rangle^{\mathsf{F}}_{\theta}$
    - $\langle \xi, \eta \rangle^{\mathsf{F}}_{\theta} = -\mathbb{E}_{\boldsymbol{x}}[\mathrm{D}^2\, L_{\boldsymbol{x}}(\theta)[\xi, \eta]]$,
      Fisher metric
    - $\{e_i\}$ orthonormal basis of $T_\theta\mathcal{M}$
      associated to $\langle\cdot,\cdot\rangle^{\mathcal{M}}_{\theta}$



$\Rightarrow$ allows to exhibit different properties from the classical case

- $\boldsymbol{x} \sim \mathrm{CES}(0, \boldsymbol{R})$, $\boldsymbol{R} \in \mathcal{H}_p^{++}$, with log-likelihood

$$L_{\boldsymbol{x}}^{++}(\boldsymbol{R}) = \log(g(\boldsymbol{x}^H \boldsymbol{R}^{-1} \boldsymbol{x})) - \log\det(\boldsymbol{R})^{-1} \qquad g : \mathbb{R}^+ \to \mathbb{R}^+$$

e.g. gaussian ($g(t) = \exp(-t)$), generalized gaussian, Student t

- Fisher metric of elliptical distributions on $\mathcal{H}_p^{++}$: [Breloy et al., 2018]

$$\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle_{\boldsymbol{R}}^{\mathrm{F},++} = \alpha^{\mathrm{F}} \operatorname{trace}(\boldsymbol{R}^{-1} \boldsymbol{\xi} \boldsymbol{R}^{-1} \boldsymbol{\eta}) + \beta^{\mathrm{F}} \operatorname{trace}(\boldsymbol{R}^{-1} \boldsymbol{\xi}) \operatorname{trace}(\boldsymbol{R}^{-1} \boldsymbol{\eta})$$

- low-rank structure – few data [Sun et al., 2016, Bouchard et al., 2020]

$$\boldsymbol{R} = \boldsymbol{I}_p + \boldsymbol{H} \qquad \boldsymbol{H} \in \mathcal{H}_{p,k}^+$$

$\Rightarrow$ geometry of $\mathcal{H}_{p,k}^+$: several possibilities, none ideal

e.g. [Bonnabel and Sepulchre, 2009, Vandereycken et al., 2012, Massart and Absil, 2018]

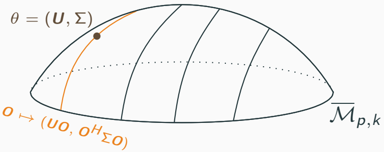
Geometry of low-rank structured covariance matrices

- $\boldsymbol{H} = \overline{\varphi}(\boldsymbol{U}, \Sigma) = \boldsymbol{U}\Sigma\boldsymbol{U}^H$ with $(\boldsymbol{U}, \Sigma) \in \overline{M}_{p,k} = \mathrm{St}_{p,k} \times \mathcal{H}_k^{++}$
- $\forall \boldsymbol{O} \in \mathcal{U}_k, \quad \overline{\varphi}(\boldsymbol{U}\boldsymbol{O}, \boldsymbol{O}^H\Sigma\boldsymbol{O}) = \overline{\varphi}(\boldsymbol{U}, \Sigma)$

$\Rightarrow \mathcal{H}_{p,k}^{+}$ isomorphic to quotient [Bonnabel and Sepulchre, 2009, Bouchard et al., 2020]

$$\mathcal{M}_{p,k} = \overline{\mathcal{M}}_{p,k}/\mathcal{U}_k = \{\pi(\boldsymbol{U}, \Sigma) : \quad (\boldsymbol{U}, \Sigma) \in \overline{M}_{p,k}\}$$

$$\text{où} \quad \pi(\boldsymbol{U}, \Sigma) = \{(\boldsymbol{U}\boldsymbol{O}, \boldsymbol{O}^H\Sigma\boldsymbol{O}) : \quad \boldsymbol{O} \in \mathcal{U}_k\}$$



quotient $\mathcal{M}_{p,k}$: Riemannian geometry

- $\theta = (\boldsymbol{U}, \Sigma) \in \overline{\mathcal{M}}_{p,k}$
- $T_\theta \overline{\mathcal{M}}_{p,k} = \{(\boldsymbol{U}\Omega_\xi + \boldsymbol{U}_\perp \boldsymbol{K}_\xi, \boldsymbol{\xi}_\Sigma) : \ \Omega_\xi \in \mathcal{A}_k, \ \boldsymbol{K}_\xi \in \mathbb{R}^{(p-k)\times k}, \ \boldsymbol{\xi}_\Sigma \in \mathcal{S}_k\}$

$\boldsymbol{U}_\perp \in \mathrm{St}_{p,(p-k)}$ such that $\boldsymbol{U}^H \boldsymbol{U}_\perp = 0$

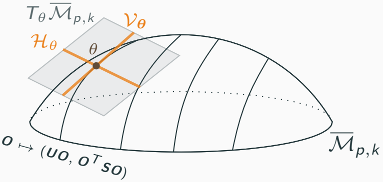
$$\langle \xi, \eta \rangle_\theta = \frac{1}{2} \operatorname{trace}(\Omega_\xi^H \Omega_\eta) + \operatorname{trace}(\boldsymbol{K}_\xi^H \boldsymbol{K}_\eta) + \alpha \operatorname{trace}(\Sigma^{-1} \boldsymbol{\xi}_\Sigma \Sigma^{-1} \boldsymbol{\eta}_\Sigma) + \beta \operatorname{trace}(\Sigma^{-1} \boldsymbol{\xi}_\Sigma) \operatorname{trace}(\Sigma^{-1} \boldsymbol{\eta}_\Sigma)$$

- vertical space – tangent space to the equivalence class

$$\mathcal{V}_\theta = \{(\boldsymbol{U}\Omega, \Sigma\Omega - \Omega\Sigma) : \ \Omega \in \mathcal{A}_k\}$$

- horizontal space – orthogonal complement to $\mathcal{V}_\theta$ according to $\langle \cdot, \cdot \rangle_\theta$

$$\mathcal{H}_\theta = \{\xi \in T_\theta \overline{\mathcal{M}}_{p,k} : \ \Omega_\xi = 2\alpha(\Sigma^{-1} \boldsymbol{\xi}_\Sigma - \boldsymbol{\xi}_\Sigma \Sigma^{-1})\}$$



quotient $\mathcal{M}_{p,k}$: error measure

- Riemannian distance not known $\Rightarrow$ alternative error measure
- alternative horizontal space – complement to $\mathcal{V}_\theta$, non orthogonal

[Bonnabel and Sepulchre, 2009]

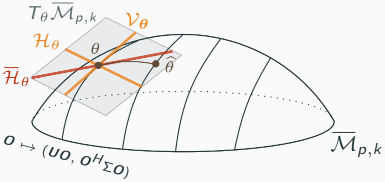
$$\begin{aligned}\overline{\mathcal{H}}_\theta &= \{\xi \in T_\theta \overline{\mathcal{M}}_{p,k} : \Omega_\xi = 0\} \\ &= \{(\boldsymbol{U}_\perp \boldsymbol{K}_\xi, \boldsymbol{\xi}_\Sigma) : \boldsymbol{K}_\xi \in \mathbb{R}^{(p-k)\times k}, \boldsymbol{\xi}_\Sigma \in \mathcal{S}_k\}\end{aligned}$$

- induces divergence onto $\mathcal{M}_{p,k}$

$$d(\pi(\theta), \pi(\hat{\theta})) = \|\Theta\|_2^2 + \alpha \left\| \log(\Sigma^{-1/2} \boldsymbol{O} \hat{\boldsymbol{O}}^H \hat{\Sigma} \hat{\boldsymbol{O}} \boldsymbol{O}^H \Sigma^{-1/2}) \right\|_2^2 + \beta \left(\log \det(\Sigma^{-1} \hat{\Sigma}) \right)^2$$

$$\boldsymbol{U}^T \hat{\boldsymbol{U}} = \boldsymbol{O} \cos(\Theta) \hat{\boldsymbol{O}}^T$$

$$\Rightarrow \mathrm{err}(\theta, \hat{\theta}) = d(\theta, \hat{\theta})$$



quotient $\mathcal{M}_{p,k}$: performance bound

$$\mathbb{E}_{\hat{\theta}}[\mathrm{err}(\theta, \hat{\theta})] = \mathbb{E}_{\hat{\theta}}[d(\theta, \hat{\theta})] \geq \mathrm{trace}(\bar{\boldsymbol{F}}^{-1})$$

$$F_{ij} = \langle e_i, e_j \rangle_{\theta}^{\mathsf{F}, \overline{\mathcal{M}}_{p,k}}$$

- $\{e_i\}$, orthonormal basis of $\overline{\mathcal{H}}_\theta$ according to $\langle \cdot, \cdot \rangle_\theta$:

$$\Big\{ \{(\boldsymbol{U}_\perp \boldsymbol{K}^{ij}, 0), (i \boldsymbol{U}_\perp \boldsymbol{K}^{ij}, 0)\}_{\substack{1 \leq i \leq p-k, \\ 1 \leq j \leq k}}, \quad \{(0, \tfrac{1}{\sqrt{\alpha}} \boldsymbol{\Sigma}^{1/2} \boldsymbol{H}^{ij} \boldsymbol{\Sigma}^{1/2} + \tfrac{\sqrt{\alpha} - \sqrt{\alpha + k\beta}}{k\sqrt{\alpha}\sqrt{\alpha + k\beta}} \mathrm{trace}(\boldsymbol{H}^{ij}) \boldsymbol{\Sigma})\}_{1 \leq j \leq i \leq k}, \quad \{(0, \tfrac{1}{\sqrt{\alpha}} \boldsymbol{\Sigma}^{1/2} \widetilde{\boldsymbol{H}}^{ij} \boldsymbol{\Sigma}^{1/2})\}_{1 \leq j < i \leq k} \Big\}$$

  - $\boldsymbol{K}^{ij} \in \mathbb{R}^{(p-k) \times k}$: element ij equals 1, 0 elsewhere
  - $\boldsymbol{H}^{ii} \in \mathbb{R}^{k \times k}$: element ii equals 1, 0 elsewhere
  - $\boldsymbol{H}^{ij} \in \mathbb{R}^{k \times k}$, $i \neq j$: elements ij and ji equal $1/\sqrt{2}$, 0 elsewhere
  - $\widetilde{\boldsymbol{H}}^{ij} \in i\mathbb{R}^{k \times k}$, $i \neq j$: elements ij and ji equal $i/\sqrt{2}$ et $-i/\sqrt{2}$, 0 elsewhere

- $\langle \cdot, \cdot \rangle_{\cdot}^{\mathsf{F}, \overline{\mathcal{M}}_{p,k}}$: Fisher metric onto $\overline{\mathcal{M}}_{p,k}$

quotient $\mathcal{M}_{p,k}$: Fisher metric of elliptical distributions

- $\overline{\varphi}(\theta) = \boldsymbol{U}\Sigma\boldsymbol{U}^H$ $\qquad\qquad$ $\mathrm{D}\,\overline{\varphi}(\theta)[\xi] = \boldsymbol{U}\boldsymbol{\xi}_\Sigma\boldsymbol{U}^H + \boldsymbol{\xi}_{\boldsymbol{U}}\Sigma\boldsymbol{U}^H + \boldsymbol{U}\Sigma\boldsymbol{\xi}_{\boldsymbol{U}}$

- log-likelihood: $\qquad L_{\boldsymbol{x}}^{\overline{\mathcal{M}}_{p,k}}(\theta) = L_{\boldsymbol{x}}^{++}(\boldsymbol{I}_p + \overline{\varphi}(\theta))$

- Fisher metric:

$$\langle \xi, \eta \rangle_\theta^{\mathsf{F},\overline{\mathcal{M}}_{p,k}} = \langle \mathrm{D}\,\overline{\varphi}(\theta)[\xi], \mathrm{D}\,\overline{\varphi}(\theta)[\eta] \rangle_{\boldsymbol{I}_p+\overline{\varphi}(\theta)}^{\mathsf{F},++}$$

  where

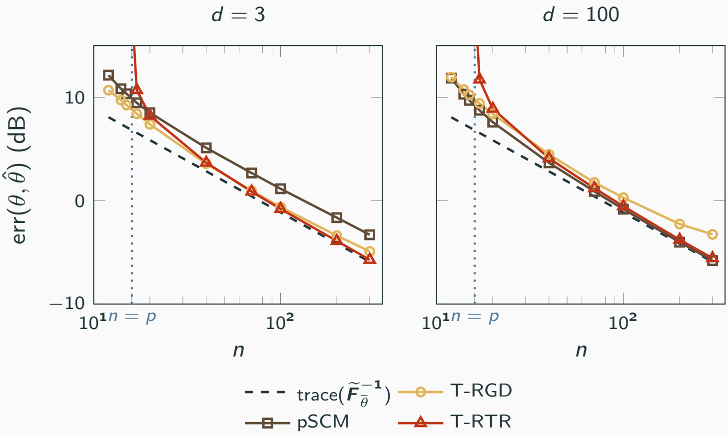
$$\langle \boldsymbol{\xi}_{\boldsymbol{R}}, \boldsymbol{\eta}_{\boldsymbol{R}} \rangle_{\boldsymbol{R}}^{\mathsf{F},++} = \alpha^{\mathsf{F}} \operatorname{trace}(\boldsymbol{R}^{-1}\boldsymbol{\xi}_{\boldsymbol{R}}\boldsymbol{R}^{-1}\boldsymbol{\eta}_{\boldsymbol{R}}) + \beta^{\mathsf{F}} \operatorname{trace}(\boldsymbol{R}^{-1}\boldsymbol{\xi}_{\boldsymbol{R}}) \operatorname{trace}(\boldsymbol{R}^{-1}\boldsymbol{\eta}_{\boldsymbol{R}})$$

Numerical illustrations

- covariance matrix: $\boldsymbol{R} = \boldsymbol{I}_p + \sigma \boldsymbol{U} \Sigma \boldsymbol{U}^H$
  - $p = 16,\ k = 8$
  - $\boldsymbol{U} \in \mathrm{St}_{p,k}$: random orthgonal matrix
  - $\Sigma \in \mathcal{H}_k^{++}$: diagonal matrix,
    minimal / maximal element: $1/\sqrt{c}$, $\sqrt{c}$ (c=20: conditionning)
    other elements: random, uniform distribution between $1/\sqrt{c}$ and $\sqrt{c}$
    trace normalisée : $\mathrm{trace}(\Sigma) = \mathrm{trace}(\boldsymbol{I}_k) = k$
  - $\sigma = 50$: free parameter
- data :
  - 500 sets of $n \in [12, 300]$ random samples
  - t-distribution, $d = 3$ (non gaussian) and $d = 100$ (almost gaussian) degrees of freedom

[Ollila et al., 2012]

Numerical illustrations



Conclusions and perspectives

Conclusions and perspectives

- Riemannian geometry:
  - powerful tool for signal processing and machine learning
  - exploits intrinsic structure of data and parameters
  - quite complete and allows modularity – modeling, optimization, performance analysis

- when to use Riemannian geometry and optimization ?
  - data and/or parameters possess a structure: constraints, invariances (geometrical properties of model)
  - metric of particular interest with respect to model (*e.g.*, information theory)

- issue: for some manifolds, geometry complicated to fully characterize

  e.g., low rank symmetric positive semidefinite matrices

$\Rightarrow$ go beyond Riemannian geometry to provide simple and efficient geometrical objects

Absil, P.-A., Mahony, R., and Sepulchre, R. (2008).
Optimization Algorithms on Matrix Manifolds.
Princeton University Press, Princeton, NJ, USA.

Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. (2007).
Geometric means in a novel vector space structure on symmetric positive-definite matrices.
***SIAM journal on matrix analysis and applications*, 29(1):328–347.**

Bhatia, R. (2009).
Positive definite matrices.
Princeton University Press.

Bhatia, R., Jain, T., and Lim, Y. (2017).
On the Bures-Wasserstein distance between positive definite matrices.
***preprint*.**

Bonnabel, S. and Sepulchre, R. (2009).
Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank.
***SIAM Journal on Matrix Analysis and Applications*, 31(3):1055–1070.**

Bouchard, F., Breloy, A., Ginolhac, G., Renaux, A., and Pascal, F. (2020).
A Riemannian framework for low-rank structured elliptical models.
***IEEE Transactions on Signal Processing*, submitted. Available on arxiv.**

Boumal, N. (2013).
On intrinsic Cramér-Rao bounds for Riemannian submanifolds and quotient manifolds.
***IEEE Transactions on Signal Processing*, 61(7):1809–1821.**

Breloy, A., Ginolhac, G., Renaux, A., and Bouchard, F. (2018).
Intrinsic Cramér–Rao bounds for scatter and shape matrices estimation in CES distributions.
***IEEE Signal Processing Letters*, 26(2):262–266.**

Cardoso, J.-F. and Souloumiac, A. (1993).
Blind beamforming for non Gaussian signals.
***IEEE Proceedings-F*, 140(6):362–370.**

Chebbi, Z. and Moakher, M. (2012).
Means of Hermitian positive-definite matrices based on the log-determinant α-divergence function.
***Linear Algebra and its Applications*, 436(7):1872–1889.**

Comon, P. and Jutten, C. (2010).
Handbook of Blind Source Separation: Independent Component Analysis and Applications.
Academic Press, 1st edition.

Massart, E. and Absil, P.-A. (2018).
Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices.
***Technical Report UCL-INMA-2018.06*.**

Moreau, E. and Macchi, O. (1994).
A one stage self-adaptive algorithm for source separation.
In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1994. ICASSP-94.*, volume 3, pages 49–52.

Ollila, E., Tyler, D. E., Koivunen, V., and Poor, H. V. (2012).
Complex elliptically symmetric distributions: Survey, new results and applications.
***IEEE Transactions on Signal Processing*, 60(11):5597–5625.**

Pham, D.-T. (2000).
Joint approximate diagonalization of positive definite Hermitian matrices.
***SIAM J. Matrix Anal. Appl.*, 22(4):1136–1152.**

Skovgaard, L. T. (1984).
A Riemannian geometry of the multivariate normal model.
***Scandinavian Journal of Statistics*, pages 211–223.**

Smith, S. T. (2005).
Covariance, subspace, and intrinsic Cramér-Rao bounds.
***IEEE Transactions on Signal Processing*, 53(5):1610–1630.**

Sra, S. (2013).
Positive definite matrices and the s-divergence.
***arXiv preprint arXiv:1110.1773*.**

Sun, Y., Babu, P., and Palomar, D. P. (2016).
Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions.
***IEEE Transactions on Signal Processing*, 64(14):3576–3590.**

Vandereycken, B., Absil, P.-A., and Vandewalle, S. (2012).
A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank.
***IMA Journal of Numerical Analysis*, 33(2):481–514.**

Villani, C. (2008).
Optimal transport: old and new, volume 338.
Springer Science & Business Media.