Information geometry in elliptical distributions

Part II of "Riemannian and information geometry in signal processing and machine learning," EUSIPCO 2022

Florent Bouchard, Arnaud Breloy and Ammar Mian

Statistics in signal processing and machine learning

Statistical point of view is ubiquitous:

- **Data** appears as the result of a **random processes** (uncertainties)
- Cast **statistical models** that reasonably fit **empirical histograms**
- Derive **processes** that achieve certain **average performance** for a **task**

(fitting, estimation, detection, classification, prediction)

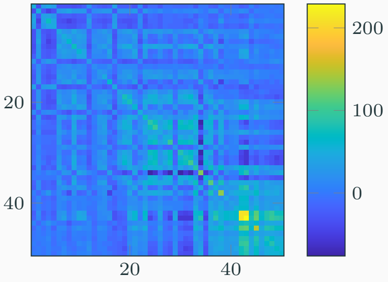
L. L. Scharf, C. Demeure, "Statistical signal processing: detection, estimation, and time series analysis," Prentice Hall, 1991

T. Hastie, R. Tibshirani, J. Friedman, "The Elements of Statistical Learning," Springer-Verlag, 2009

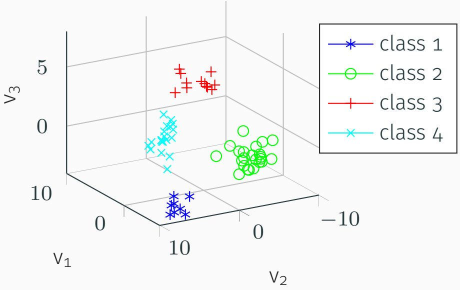
Parametric approach

Represent or analyze the data $\mathbf{x}$ through some statistical 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 re-define Riemannian tools
- **Analyze**
  - Information geometry
  - Intrinsic Cramér-Rao bounds
- **Solve**
  - Riemannian optimization and geodesic convexity
  - Examples where numerical stability is improved
- **Apply**
  - Change detection in satellite image time series

Motivation for elliptical distributions

Objective: find a model $f(\mathbf{x}, \theta)$

- $\mathbf{x}$ is a **sample** in $\mathbb{R}^p$ or $\mathbb{C}^p$ (unstructured)
- f is a **pdf**
- θ **parameterizes** the pdf

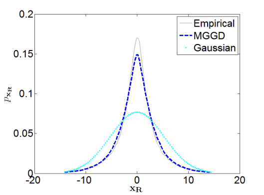
Challenges from real data:

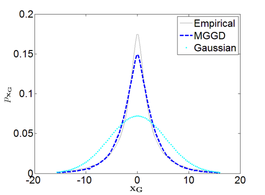
- **Non-Gaussian**, **heavy-tailed** distributions
- **Outliers**

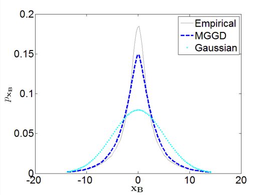
Elliptical models good entry point for this tutorial =)

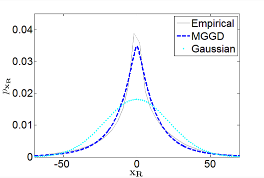
- **Large family** that that generalizes the multivariate Gaussian distribution
- Still parameterized through **1st and 2nd order moments** (mean, covariance)
- **Better fit** to empirical histograms $\rightarrow$ **better results**

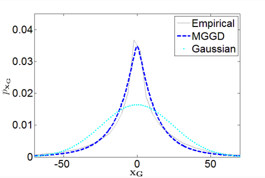
Motivating real-data examples (1/2)

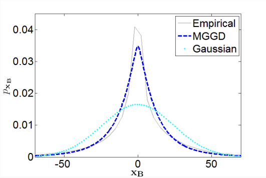








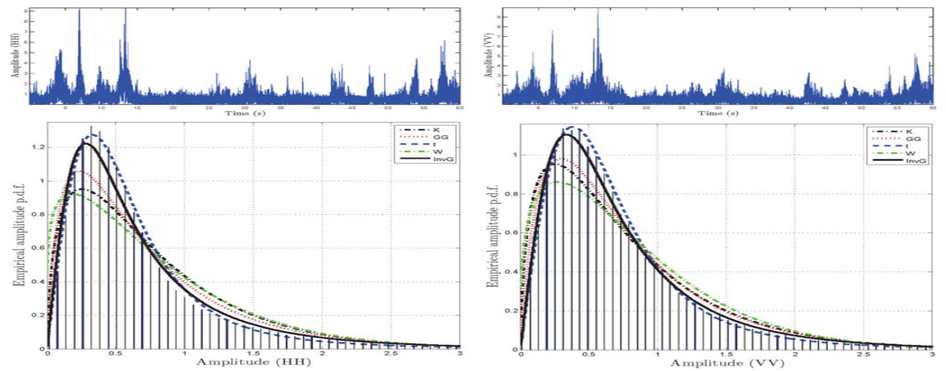




Bark.0000 and Leaves.0008 from VisTex and marginal distributions of wavelet coefficients from RGB channels.

F. Pascal, L. Bombrun, J-Y. Tourneret, Y. Berthoumieu, “Parameter estimation for multivariate generalized Gaussian distributions,” IEEE TSP, 2013

Motivating real-data examples (2/2)



Modulus of HH and VV band of Shore of Lake Ontario sensed by McMaster IPIX radar

E. Ollila, D. E. Tyler, V. Koivunen, H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," IEEE TSP, 2012

Elliptical models

Complex elliptically symmetric distributions (CES)

$\mathbf{x} \sim \mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ if its pdf can be written

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

where $g : [0, \infty) \to [0, \infty)$ is the **density generator** and

- $\boldsymbol{\mu} \in \mathbb{C}^p$ is the symmetry **center**
- $\boldsymbol{\Sigma} \in \mathcal{H}_p^{++}$ is the **scatter matrix**

If $\mathbf{x}$ has finite 2^{nd}-order moment, the **covariance matrix** is $\mathbb{E}\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^H\right] = \alpha\, \boldsymbol{\Sigma}$

- $\alpha = -2\varphi'(0)$,
- φ is defined by the characteristic function $c_{\mathbf{x}}(\mathbf{t}) = \exp(i\mathbf{t}^H \boldsymbol{\mu})\, \varphi(\mathbf{t}^H \boldsymbol{\Sigma} \mathbf{t})$

Practical CES representation

Stochastic representation theorem

$\mathbf{x} \sim \mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ iif it admits the stochastic representation

$$\mathbf{x} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{\mathcal{Q}}\boldsymbol{\Sigma}^{1/2}\mathbf{u}$$

where

- $\mathbf{u} \sim \mathcal{U}\left(\mathbb{C}\mathcal{S}^p\right)$ follow an uniform distribution on unit complex p-sphere
- $\mathcal{Q}$ is the 2^{nd}**-order modular variate**, independent of $\mathbf{u}$, with pdf

$$p(\mathcal{Q}) = \delta_{p,g}^{-1}\mathcal{Q}^{p-1}g(\mathcal{Q})$$

Interpretation:

- $\boldsymbol{\Sigma}$ pilots the **shape of the ellipsoid** (privileged direction)
- $\mathcal{Q}$ (equivalently g) models **amplitude fluctuations** (possibly heavy tails)

Some remarks on CES properties

1. **One-to-one relation** between pdf of $\mathcal{Q}$ and g
2. **Ambiguity**: $(\mathcal{Q}, \mathbf{\Sigma})$ and $(c^{-1}\mathcal{Q}, c\mathbf{\Sigma})$, $c > 0$ are valid stochastic representations of $\mathbf{x}$
   $\Rightarrow$ requires normalization constraint
3. **Covariance matrix**: $\mathbb{E}\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^H\right] = \mathbb{E}[\mathcal{Q}]\mathbf{\Sigma}/p$, if $\mathbb{E}[\mathcal{Q}]$ exists
4. **Random number generation**:
   - Draw a 2^{nd}-order modular variate $\mathcal{Q}$ from its pdf $p()$
   - Draw $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_p)$, then $\mathbf{u} \overset{d}{=} \mathbf{n}/|\mathbf{n}|\ \mathcal{U} \sim (\mathbb{C}S^p)$
   - Set $\mathbf{x} \overset{d}{=} \boldsymbol{\mu} + \sqrt{\mathcal{Q}}\mathbf{\Sigma}^{1/2}\mathbf{u}$

Important related distribution families

Compound Gaussian (CG) aka spherically invariant random vectors (SIRV)

$\mathbf{x} \sim \mathcal{CG}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f_\tau)$ iif it admits the stochastic CG-representation

$$\mathbf{x} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{\tau}\mathbf{n}$$

where

- $\tau \geq 0$ is called the **texture**, with pdf f_τ that is independent of $\mathbf{n}$
- $\mathbf{n} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is called the **speckle**.

Note: subclass of CES because if $\mathbf{n}_0 \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{I})$, then $\mathbf{n}_0 \stackrel{d}{=} \sqrt{s}\mathbf{u}$ with $s \sim \Gamma(1, p)$

Mixture of scaled Gaussian distributions (MSG)

$\mathbf{x}_i \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \tau_i\boldsymbol{\Sigma})$, where τ_i is unknown deterministic

Main examples (1/2)

Multivariate Gaussian distribution

CG: $f_\tau = \delta_1$ (or CES with $\mathcal{Q} \sim \Gamma(1, p)$)

Multivariate t-distribution with degree of freedom ν

CG: $\tau^{-1} \sim \Gamma(\nu/2, 2/\nu)$, where $\nu > 0$

- Encompasses Complex Cauchy dist. ($\nu = 1$) and CN dist. ($\nu \to \infty$)
- Finite 2nd-order moment for $\nu > 2$

K-distribution with shape parameter ν

CG: $\tau \sim \Gamma(\nu, 1/\nu)$, where $\nu > 0$

- Encompasses heavy-tailed dist. ($\nu \downarrow$) and CN dist. ($\nu \to \infty$)
- $\mathbb{E}[\tau] = 1 \implies \boldsymbol{\Sigma} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right]$

Main examples (2/2)

GG distribution with parameters s and η

- CES: $\mathcal{Q} =_d G^{1/s}$ where $G \sim \Gamma(m/s, \eta)$, $s, \eta > 0$
- PDF: $f_{\mathbf{x}}(\mathbf{x}) = cte\,|\mathbf{\Sigma}|^{-1} \exp\left(-(\eta\, \mathbf{x}^H \mathbf{\Sigma}^{-1} \mathbf{x})^s\right)$

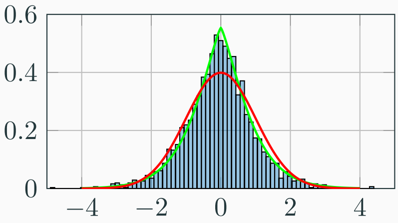
- Complex analog of the exponential power family, also called Box-Tiao distributions
- Subclass of multivariate symmetric Kotz-type distributions
- Case $s = 1 \implies$ CN dist.
- Heavier tailed than normal for $s < 1$ and lighter tailed for $s > 1$
- $s = 1/2 \implies$ generalization of Laplace dist.

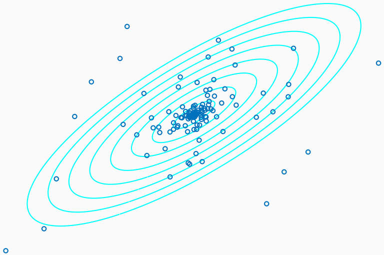
Wrapping-up

Complex elliptically symmetric distributions (CES)

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

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





Pointers and keywords

On CES

K. T. Fang, Y. T. Zhang, "Generalized Multivariate Analysis," Springer Verlag, 1990

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

On non-circularity

P. J. Schreier, L. L. Scharf, "Statistical signal processing of complex-valued data: the theory of improper and noncircular signals," Cambridge university press, 2010

H. Abeida, J-P. Delmas "Slepian–Bangs formula and Cramér–Rao bound for circular and non-circular complex elliptical symmetric distributions," IEEE Signal Processing Letters, 26(10), 1561-1565, 2019

Distributions on manifolds

K. V. Mardia, P. E. Jupp, "Directional statistics,"New York: Wiley, 2000

X. Pennec, "Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements," NSIP (Vol. 3, pp. 194-198), 1999

S. Said, L. Bombrun, Y. Berthoumieu, J. H. Manton, "Riemannian Gaussian distributions on the space of symmetric positive definite matrices," IEEE Transactions on Information Theory, 63(4), 2153-2170, 2017

Outline

- **Design**
  - Examples of f and $\boldsymbol{\theta}$ from elliptical distributions
  - Remark that $\boldsymbol{\theta} \in \mathcal{M} \implies$ pretext to re-define Riemannian tools
- **Analyze**
  - Information geometry
  - Intrinsic Cramér-Rao bounds
- **Solve**
  - Riemannian optimization and geodesic convexity
  - Examples where numerical stability is improved
- **Apply**
  - Change detection in satellite image time series

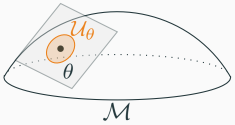
Structured parameter space as a manifold

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

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

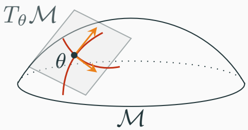
turn out to be a **manifold** $\mathcal{M}$ (locally diffeomorphic to $\mathbb{R}^d$, with $\dim(\mathcal{M}) = d$)

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



Riemannian manifolds (1/2)

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

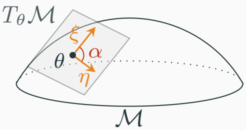


- Curve $\gamma : \mathbb{R} \to \mathcal{M}$, $\gamma(0) = \theta$
- Derivative: $\dot{\gamma}(0) = \lim_{t\to 0} \frac{\gamma(t) - \gamma(0)}{t}$
- Tangent space $T_\theta\mathcal{M} = \{\dot{\gamma}(0) : \ \gamma : \mathbb{R} \to \mathcal{M}, \ \gamma(0) = \theta\}$

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 $\mathcal{M}$

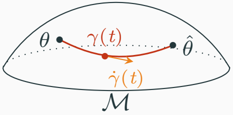
Geodesics $\gamma : [0, 1] \to \mathcal{M}$

- 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 re-define Riemannian tools
- **Analyze**
  - Information geometry
  - Intrinsic Cramér-Rao bounds
- **Solve**
  - Riemannian optimization and geodesic convexity
  - Examples where numerical stability is improved
- **Apply**
  - Change detection in satellite image time series

Fisher information metric

Sample set $\{\mathbf{x}_k\}$, iid according to the distribution $f(\mathbf{x};\theta)$, with $\theta \in \mathcal{M}$ (smooth manifold)

Score vector $s_\theta(\{\mathbf{x}_i\}) = \nabla_\theta \ln f(\{\mathbf{x}_i\};\theta)$

The **Fisher information metric** $\langle \cdot, \cdot \rangle_\theta^{\mathrm{FIM}}$ is the **covariance matrix of the score vector**

In practice

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

and **polarization formula** $\langle \xi_i, \xi_j \rangle_\theta^{\mathrm{FIM}} = \frac{1}{4}(\langle \xi_i + \xi_j, \xi_i + \xi_j \rangle_\theta^{\mathrm{FIM}} - \langle \xi_i - \xi_j, \xi_i - \xi_j \rangle_\theta^{\mathrm{FIM}})$

Fisher-Rao geometry

The **FIM** $\langle \cdot, \cdot \rangle_\theta^{\mathrm{FIM}}$ defines a **Riemannian metric** on $T_\theta \mathcal{M}$

$\mathcal{M}$ equipped with $\langle \cdot, \cdot \rangle_\theta^{\mathrm{FIM}}$ is called a **Fisher-Rao manifold**

- **a geometry** for $\mathcal{M}$ (Levi-Civita connexion, geodesics, distances, ...)
- **an implicit a geometry** for a family of statistical models

$$\mathrm{dist}^2_{\mathrm{Rao}}(f(\cdot, \theta_1), f(\cdot, \theta_2)) = \mathrm{dist}^2_{\mathrm{FIM}}(\theta_1, \theta_2)$$

with many possible applications!

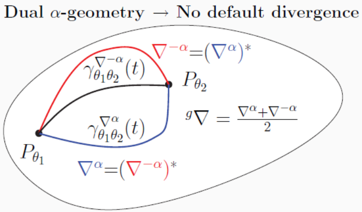
More generally, **information geometry** studies Riemannian manifolds whose points correspond to probability distributions

Information geometry is a broad field

Fisher-Rao: FIM paired with its Levi-Civita connection

Chentsov: $\mathcal{M}$ equipped with more general connections

Amari: α-geometry (dual affine connections coupled to the FIM)



Other generalizations

- non parametric, semi-parametric, ...
- Interactions between information geometry and optimal transport

(links between Wasserstein distances and divergences, geometry induced by Wasserstein metric, ...)

Many results such as **generalized divergences** and tools for **Riemannian optimization**

F. Nielsen, "The many faces of information geometry," Not. Am. Math., 2022

Example: Fisher-Rao distance between multivariate Gaussian distributions

Assume $\mathbf{x} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, then

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

Taylor expansions of $\ln f$, expectation, polarization, yields

$$\langle \xi_i, \xi_j \rangle_{\boldsymbol{\Sigma}}^{\text{FIM}} = \text{Tr}\left\{\boldsymbol{\Sigma}^{-1} \xi_i \boldsymbol{\Sigma}^{-1} \xi_j\right\}$$

Conclusion the FIM for centered Gaussian models is the affine invariant metric !

The **Fisher-Rao distance between two centered $\mathbb{C}\mathcal{N}$** is then

$$\text{dist}^2_{\mathbb{C}\mathcal{N}}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = || \log \boldsymbol{\Sigma}_1^{-1/2} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1/2} ||_F^2$$

used, e.g., to compare/classify population sets $\to$ Part III

Bridges between statistics and the Riemannian geometry of $\mathcal{H}_p^{++}$! metric $\leftrightarrow$ model on $\mathbf{x}$

Some pointers

Univariate Gaussian models

S. I Costa, S. A. Santos, J. E. Strapasson, “Fisher information distance: A geometrical reading,” Discrete Applied Mathematics, 197, 59-69, 2015

Centered elliptical models

C. A. Micchelli, L. Noakes, “Rao distances,” Journal of multivariate analysis, 92(1), 97-115, 2005

General mean-covariance case unknown!

M. Calvo, J. M. Oller, “A distance between elliptical distributions based in an embedding into the Siegel group,” Journal of Computational and Applied Mathematics, 145(2), 319-334, 2002

P. S. Eriksen, “Geodesics connected with the Fischer metric on the multivariate normal manifold,” Institute of Electronic Systems, Aalborg University Centre, 1986

M. Pilté, F. Barbaresco, “Tracking quality monitoring based on information geometry and geodesic shooting,” 17th International Radar Symposium (IRS) (pp. 1-6). IEEE, 2016

Outline

- **Design**
  - Examples of f and $\boldsymbol{\theta}$ from elliptical distributions
  - Remark that $\boldsymbol{\theta} \in \mathcal{M} \implies$ pretext to re-define Riemannian tools
- **Analyze**
  - Information geometry
  - Intrinsic Cramér-Rao bounds
- **Solve**
  - Riemannian optimization and geodesic convexity
  - Examples where numerical stability is improved
- **Apply**
  - Change detection in satellite image time series

Cramér-Rao lower bound (CRLB)

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

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

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

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

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

O. Besson, Y. I. Abramovich, “On the Fisher information matrix for multivariate elliptically contoured distributions,” IEEE Signal Processing Letters, 20(11), 1130-1133, 2013

$\rightarrow$ for CES!

“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

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

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 $\mathbf{H}(\boldsymbol{\theta}) = \left.\frac{\partial \mathbf{h}(\boldsymbol{\theta})}{\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}}$

J. D. Gorman, A. O. Hero, “Lower bounds for parametric estimation with constraints,” IEEE Transactions on Information Theory, 36(6), 1285-1301, 1990

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

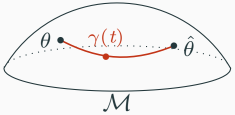
$\rightarrow$ We can turn to the framework of **intrinsic CRLB** (iCRLB)

S. T. Smith, "Covariance, subspace, and intrinsic crame/spl acute/r-rao bounds. IEEE Transactions on Signal Processing," 53(5), 1610-1630, 2005

Riemannian framework of iCRLB

Definitions:

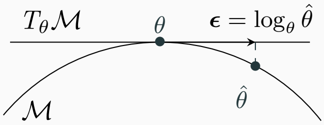
- $\theta \in \mathcal{M}$ with tangent space $T_\theta\mathcal{M}$
- $\hat{\theta} \in \mathcal{M}$ estimate of θ
- $\langle \cdot, \cdot \rangle_\theta$ chosen Riemannian metric
- $\mathrm{dist}(\cdot, \cdot)$ induced Riemannian distance
- $\{\xi_i\}$ corresponding orthonormal basis of $T_\theta\mathcal{M}$



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

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

- 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 $[\epsilon]_i = \langle \log_\theta \hat{\theta}, \xi_i \rangle_\theta$



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)

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 $\mathbb{E}\left\{\mathrm{dist}^2(\hat{\theta},\theta)\right\} \geq \mathrm{Tr}\left\{\mathbf{F}^{-1}\right\}$

"$(\cdot)^{-1}$" 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 $T_\theta\mathcal{M} \rightarrow$ useful even in the Euclidean case!

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

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

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

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

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

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

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 $\mathbf{x} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g)$, then

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

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

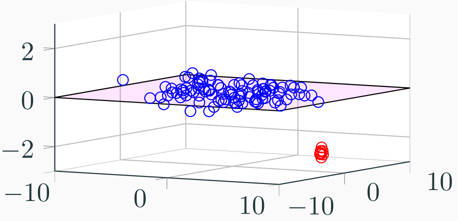
$$\mathbb{E}\left[\mathrm{dist}^2_{\mathcal{H}_p^{++}}\left(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}\right)\right] \geq \frac{1}{n}\left(\frac{p^2-1}{\alpha_g} + \frac{1}{\alpha_g(p+1)-p}\right)$$

Example 2: probabilistic PCA in CES (1/2)

Probabilistic PCA (PPCA)

$$\mathbf{x} \overset{d}{=} \mathbf{W}\mathbf{s} + \mathbf{n}$$

with $\mathbf{W} \in \mathbb{C}^{p \times k}$, $\mathbf{s} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{I}_k)$, $\mathbf{n} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$



Structured covariance matrix low-rank + identity

$$\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{\Sigma} = \mathbf{H} + \mathbf{I}, \text{ with } \mathrm{rank}(\mathbf{H}) = k$$

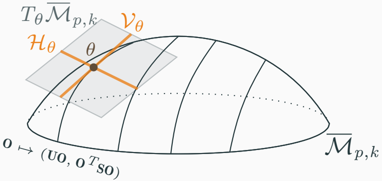
CES-PPCA generalizes the model to $\mathbf{x} \sim \mathcal{CES}(\mathbf{0}, \mathbf{H} + \mathbf{I}, g)$

M. E. Tipping, C. M Bishop, “Probabilistic principal component analysis,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(3), 611-622., 1999

Example 2: probabilistic PCA in CES (1/2)

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

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

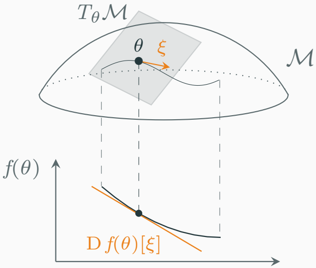
Outline

- **Design**
  - Examples of f and $\boldsymbol{\theta}$ from elliptical distributions
  - Remark that $\boldsymbol{\theta} \in \mathcal{M} \implies$ pretext to re-define Riemannian tools
- **Analyze**
  - Information geometry
  - Intrinsic Cramér-Rao bounds
- **Solve**
  - Riemannian optimization and geodesic convexity
  - Examples where numerical stability is improved
- **Apply**
  - Change detection in satellite image time series

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:

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

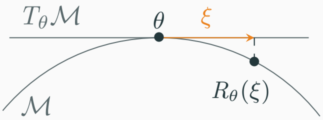
Riemannian gradient of f at θ:

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

Riemannian optimization

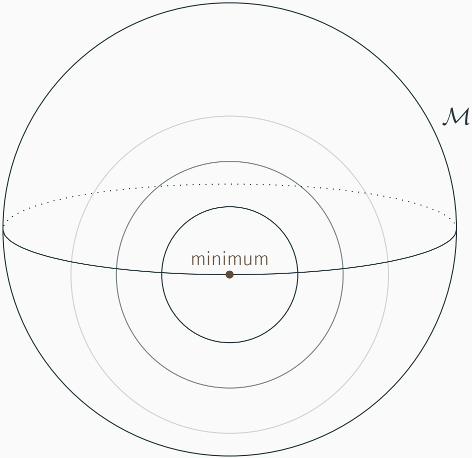
Main ingredients

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

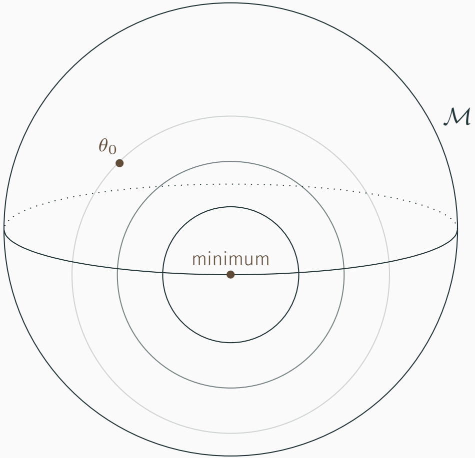


Flexibility: metric, retraction, descent method (gradient, conjugate gradient, BFGS...)

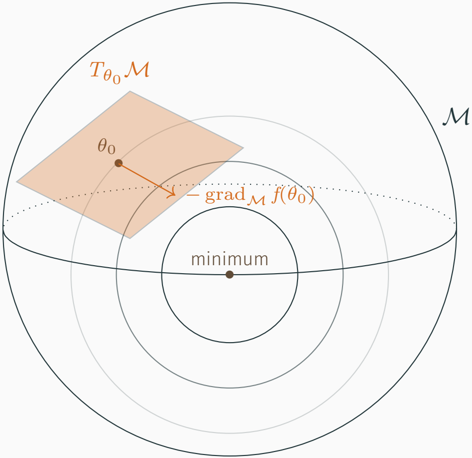
Example: Riemannian gradient descent $\theta_{i+1} = R_{\theta_i}(-t_i \, \mathrm{grad}\, f(\theta_i))$



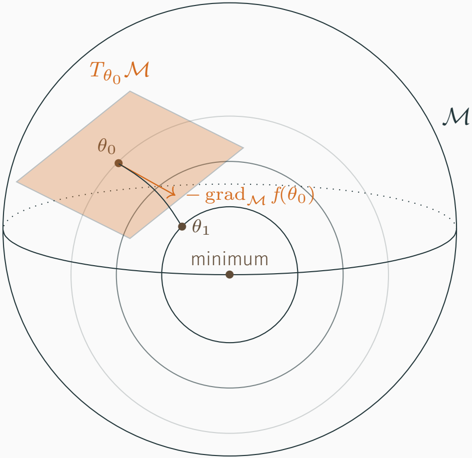
Example: Riemannian gradient descent $\theta_{i+1} = R_{\theta_i}(-t_i \, \mathrm{grad}\, f(\theta_i))$



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



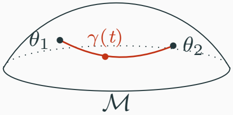
Example: Riemannian gradient descent $\theta_{i+1} = R_{\theta_i}(-t_i \,\mathrm{grad}\, f(\theta_i))$



Geodesic convexity (g-convexity)

$\mathcal{M}$ is a **g-convex set** w.r.t. geodesic $\gamma(t)$, if

$$\forall\ \theta_1, \theta_2 \in \mathcal{M}, \gamma(t) \in \mathcal{M}$$



Example: $\mathcal{H}_p^{++}$ is **g-convex** w.r.t. $\mathbf{\Sigma}(t) = \mathbf{\Sigma}_1^{1/2} \left(\mathbf{\Sigma}_1^{-1/2} \mathbf{\Sigma}_2 \mathbf{\Sigma}_1^{-1/2} \right)^t \mathbf{\Sigma}_1^{1/2}$

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

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

Property If f is **g-convex** then any local minimizer is a global minimizer on $\mathcal{M}$

Outline

- **Design**
  - Examples of f and $\boldsymbol{\theta}$ from elliptical distributions
  - Remark that $\boldsymbol{\theta} \in \mathcal{M} \implies$ pretext to re-define Riemannian tools
- **Analyze**
  - Information geometry
  - Intrinsic Cramér-Rao bounds
- **Solve**
  - Riemannian optimization and geodesic convexity
  - Examples where numerical stability is improved
- **Apply**
  - Change detection in satellite image time series

Example 1: regularized covariance matrix estimation in CES (1/3)

M-Estimators of the scatter

The minimizers of the objective function ~ CES log-likelihood

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

Satisfy the fixed-point equation

$$\boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \boldsymbol{\Sigma}^{-1}\mathbf{x}_i)\ \mathbf{x}_i\mathbf{x}_i^H \ \triangleq\ \mathcal{H}_u(\boldsymbol{\Sigma})$$

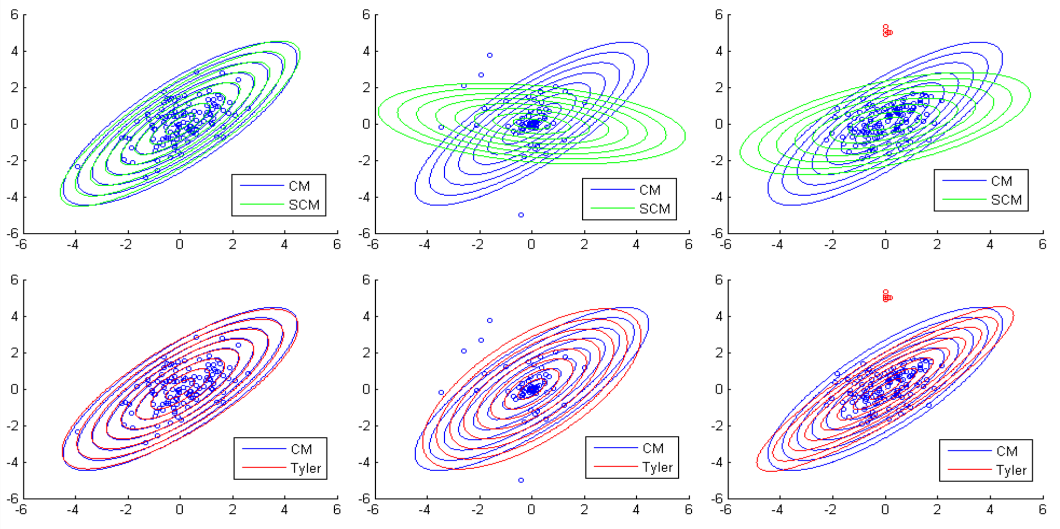
with $u = -g'(t)/g(t)$

Studied in the $70 \sim 80$'s, modern interest due to **robustness** and **new insights**

- $\mathcal{L}$ is **g-convex** following the geodesics $\boldsymbol{\Sigma}(t) = \boldsymbol{\Sigma}_1^{1/2}(\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1/2})^t\boldsymbol{\Sigma}_1^{1/2}$
- $\boldsymbol{\Sigma}_{t+1} = \mathcal{H}_u(\boldsymbol{\Sigma}_t)$ is a **majorization-minimization** algorithm

A. Wiesel, "Geodesic convexity and covariance estimation," IEEE TSP, 60(12), 6182-6189, 2012

Example 1: regularized covariance matrix estimation in CES (2/3)



Example 1: regularized covariance matrix estimation in CES (3/3)

Issue: optimality/uniqueness guaranteed, but existence requires $n > p$

Solution: regularization methods driven by g-convexity

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

Shrinkage to identity: $\mathcal{P}(\boldsymbol{\Sigma}) = \mathrm{Tr}\{\boldsymbol{\Sigma}^{-1}\}$ is g-convex

$$\boldsymbol{\Sigma}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} u(\mathbf{x}_i^H \boldsymbol{\Sigma}^{-1}(\alpha) \mathbf{x}_i)\ \mathbf{x}_i \mathbf{x}_i^H + \alpha \mathbf{I}$$

Can exist for $n < p$!

Many **generalizations** and **optimal selection** of α for various criterions

E. Ollila, D. E. Tyler, "Regularized M-estimators of scatter matrix," IEEE TSP, 62(22), 6059-6070, 2014

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

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

***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 2: robust mean and covariance estimation (2/3)

Alternatively when $\boldsymbol{\mu} = \mathbf{0}$: Tyler's estimator $\Leftrightarrow$ **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 2: robust mean and covariance estimation (3/3)

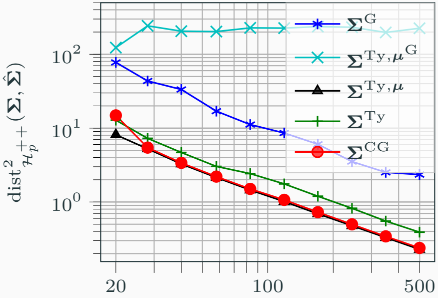
Product manifold $\mathcal{M}_{p,n} \in \mathbb{C}^p \times (\mathbb{R}_\star^+)^n \times \mathcal{SH}_p^{++}$ with ***decoupled* metric**

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

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

And resulting:

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



Example 3: robust estimator for spiked models in CES (1/2)

Spiked Tyler's estimator

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

Existing **MM algorithm**

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)

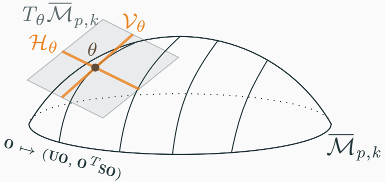
can diverge with small n

Y. Sun et al. "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," IEEE TSP, 2016

Example 3: robust estimator for spiked models in CES (2/2)

Riemannian optimization for

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



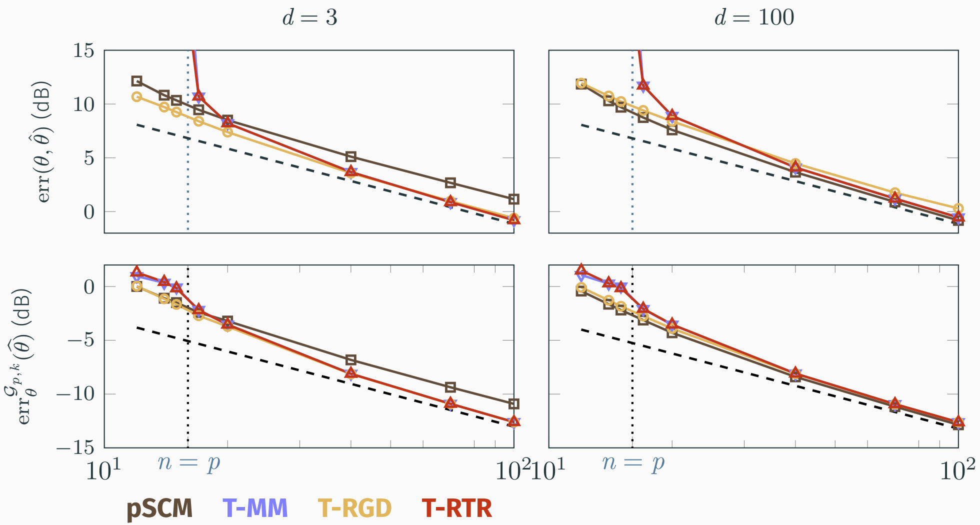
with $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H \in (\mathrm{St}(p,k) \times \mathcal{H}_k^{++})/\mathcal{U}_k$

using the metric

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

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

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

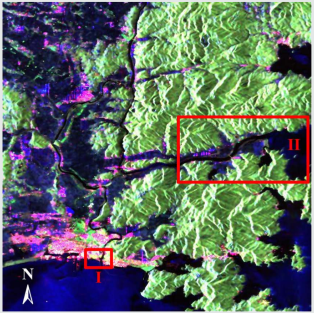


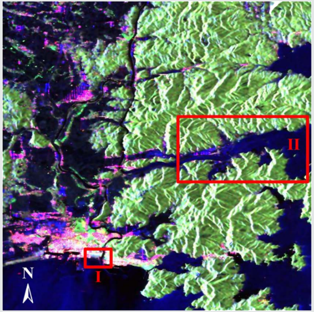
Outline

- **Design**
  - Examples of f and $\boldsymbol{\theta}$ from elliptical distributions
  - Remark that $\boldsymbol{\theta} \in \mathcal{M} \implies$ pretext to re-define Riemannian tools
- **Analyze**
  - Information geometry
  - Intrinsic Cramér-Rao bounds
- **Solve**
  - Riemannian optimization and geodesic convexity
  - Examples where numerical stability is improved
- **Apply**
  - Change detection in satellite image time series

Change detection in satellite image time-series

Monitoring natural disasters:





PolSAR images of Ishinomaki and Onagawa areas [Sato, 2012], Nov.2010 (left), Apr.2011 (right).

Problems to consider

Huge increase in the number of available acquisitions:

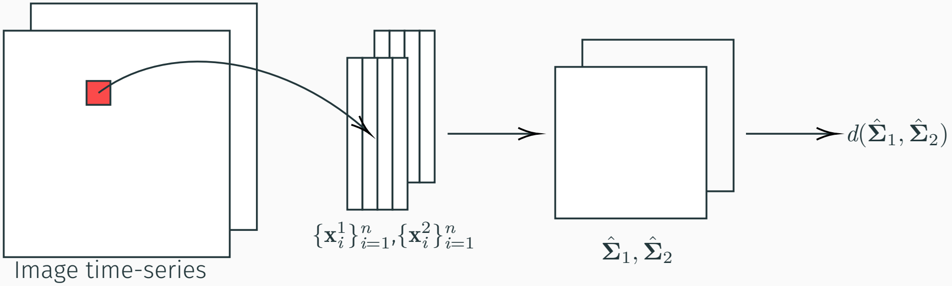
- Sentinel-1: 12 days repeat cycle, since 2014
- TerraSAR-X: 11 days repeat cycle, since 2007
- UAVSAR, ...

Detect changes

- Massive amount of data $\longrightarrow$ Automatic process
- Unlabeled data $\longrightarrow$ Unsupervised detection

Chosen approach: detection based on covariance matrix (statistical approaches)

2-step change detection



- **Covariance matrix estimation** (feature extraction)
- Evaluation of a **distance** (feature comparison)

Covariance matrix estimation

Sample covariance matrix (SCM)

Let $\{\mathbf{x}_i\}_{i=1}^n$ following $\mathbf{x} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, the ML estimate of $\boldsymbol{\Sigma}$ is

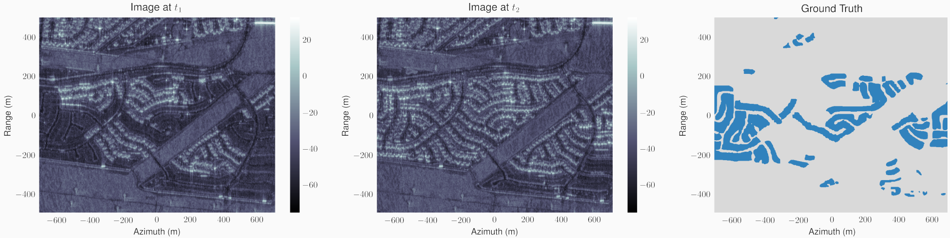
$$\hat{\boldsymbol{\Sigma}}_{\text{SCM}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^H$$

- Simple to implement
- Wishart distributed $\longrightarrow$ well established properties
- Not robust to non-Gaussian/outliers (cf. Part 2)

Distances between covariance matrices

Frobenius	$d_{\text{Fro}}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|_F^2$
Spectral Log	$d_{\text{Log}}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \|\log(\boldsymbol{\Sigma}_1) - \log(\boldsymbol{\Sigma}_2)\|_F^2$
Hotelling-Lawley	$d_{\text{HTL}}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \text{Tr}\left\{\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1}\right\}$
KL divergence	$d_{\text{KL}}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \text{Tr}\left\{\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_2\right\} + \log\left(\lvert\boldsymbol{\Sigma}_1\rvert / \lvert\boldsymbol{\Sigma}_2\rvert\right)$
Wasserstein	$d_{\text{W}}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \text{Tr}\left\{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 - 2\left(\boldsymbol{\Sigma}_2^{1/2} \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{1/2}\right)^{1/2}\right\}$
Rao	$d_{\text{Rao}}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \alpha \sum_{i=1}^{p} \log^2 \lambda_i + \beta \left(\sum_{i=1}^{p} \log \lambda_i\right)^2$ $\{\lambda_i\}_{i=1}^{p} = \text{eig}(\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_2)$

Dataset



UAVSAR SanAnd_26524_03

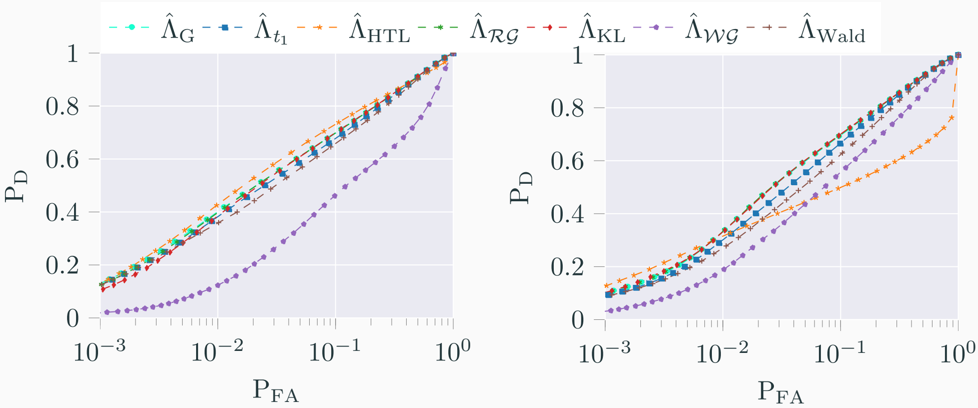
- CD between April 2009 - May 2011 [Nascimento19]
- Polarimetric data $\longrightarrow$ wavelet decomposition $\longrightarrow p = 12$ dim. pixels

A. Mian, G. Ginolhac, J-P. Ovarlez, A. Breloy, F. Pascal, "An Overview of Covariance-based Change Detection Methodologies in Multivariate SAR Image Time Series," Change Detection and Image Time Series Analysis, 2021

Compared detectors

- Plug-in detectors using SCMs ($T = 2$)
  - Λ_{HTL} Hotelling-Lawley divergence
  - Λ_{KL} KL divergence
  - $\Lambda_{\mathcal{RG}}$ Riemannian distance (Rao distance with $\alpha = 1$, $\beta = 0$)
  - $\Lambda_{\mathcal{WG}}$ Wasserstein distance
- Gaussian detection criteria
  - Λ_{G} GLRT
  - Λ_{t_1} Terrell statistic
  - Λ_{Wald} Wald statistic

Results scene 1-2 ($T = 2$)



ROC plots using a 5×5 local window for the scenes 1 and 2.

Conclusion on 2-step change detection

SCM plug-in detectors

$$\Lambda(\{\{\mathbf{x}_i^t\}_{i=1}^n\}_{t=1}^T) = d(\hat{\mathbf{\Sigma}}_{\text{SCM}}^1, \hat{\mathbf{\Sigma}}_{\text{SCM}}^2)$$

Advantages

- Practical anf flexible
- SCM is Wishart
- Various distances (invariances)
- Can change plug-in SCMs

Limitations

- $T = 2$
- CFAR: case by case study
- 2-step $\rightarrow$ "suboptimal"?
- Indirect link with $f(\mathbf{x}, \theta)$

Change detection with GLRT

Parametric probability model

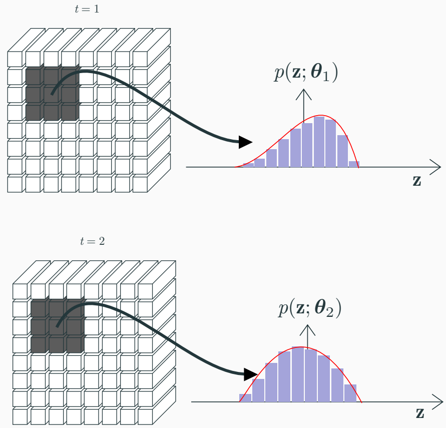
$$\mathbf{Z}_t \sim \mathcal{L}(\mathbf{Z}_t; \boldsymbol{\theta}_t)$$

Hypothesis test

$$\begin{cases} \mathrm{H}_0 : & \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \quad (\textit{no change}) \\ \mathrm{H}_1 : & \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \quad (\textit{change}) \end{cases}$$

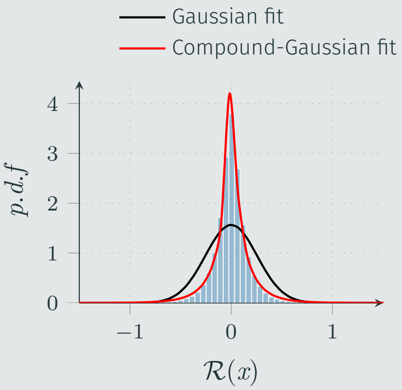
GLRT

$$\frac{\max\limits_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2} \mathcal{L}\left(\{\mathbf{Z}_1, \mathbf{Z}_2\} \; ; \; \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}\right)}{\max\limits_{\boldsymbol{\theta}_0} \mathcal{L}\left(\{\mathbf{Z}_1, \mathbf{Z}_2\} \; ; \; \boldsymbol{\theta}_0\right)} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda_{\mathrm{GLRT}}$$

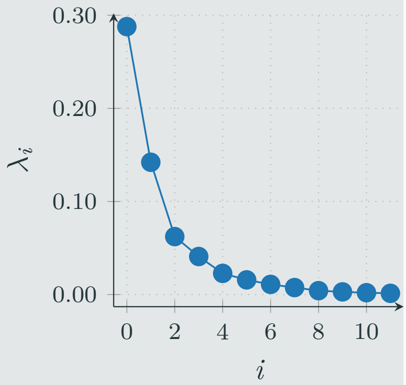


Empirical hints for the chosen model

Histogram of UAVSAR data (HH)



Spectrum of UAVSAR data (wavelets)



Covariance based change detection

Models for the GLRT in SAR-ITS: appropriate choice of $\mathcal{L}$ and θ

Gaussian

$$\mathbf{z} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$
$$\boldsymbol{\theta} = \boldsymbol{\Sigma}$$

Low-rank Gaussian

$$\mathbf{z} \sim \mathbb{C}\mathcal{N}[\mathbf{0}][\boldsymbol{\Sigma}_k + \sigma^2 \mathbf{I}]$$
$$\boldsymbol{\theta} = \boldsymbol{\Sigma}, \text{ with } \mathrm{rank}(\boldsymbol{\Sigma}_k) = k$$

Compound-Gaussian

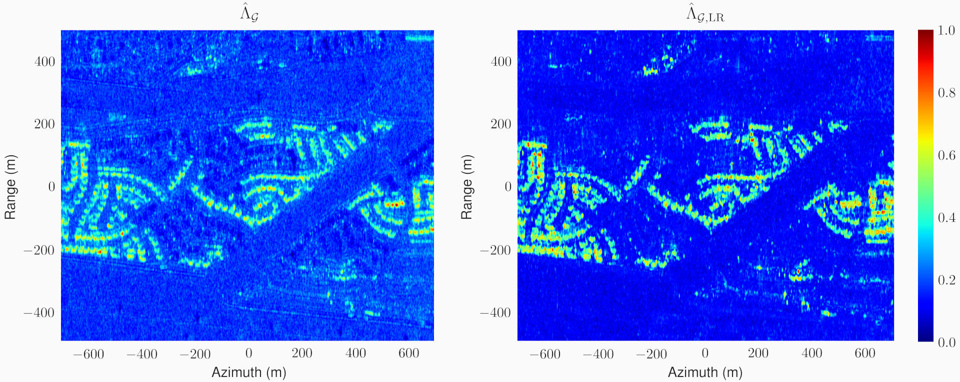
$$\mathbf{z}_i \sim \mathbb{C}\mathcal{N}[\mathbf{0}][\tau_i \boldsymbol{\Sigma}]$$
$$\boldsymbol{\theta} = \{\boldsymbol{\Sigma}, \{\tau_i\}\}$$

Low-rank Compound-Gaussian

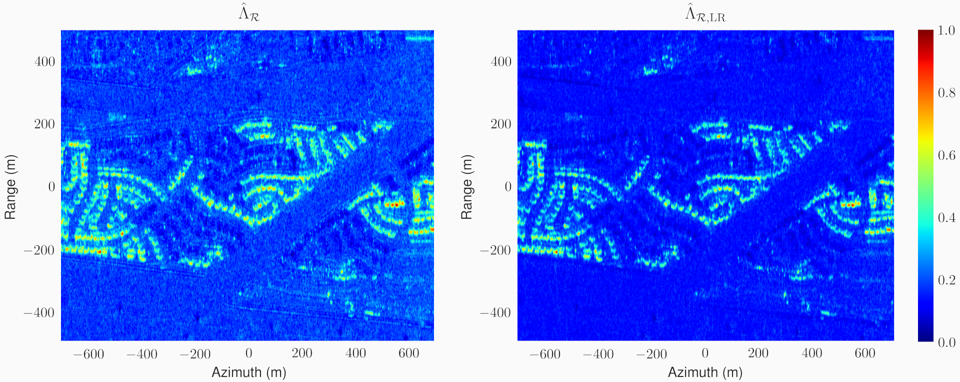
$$\mathbf{z}_i \sim \mathbb{C}\mathcal{N}[\mathbf{0}][\tau_i(\boldsymbol{\Sigma}_k + \sigma^2 \mathbf{I})]$$
$$\boldsymbol{\theta} = \{\boldsymbol{\Sigma}, \{\tau_i\}\}, \text{ with } \mathrm{rank}(\boldsymbol{\Sigma}_k) = k$$

Optimization handled with $\boldsymbol{\Sigma} = \mathbf{U}\mathbf{D}\mathbf{U}^H$ and previous techniques (Riemannian opt.)

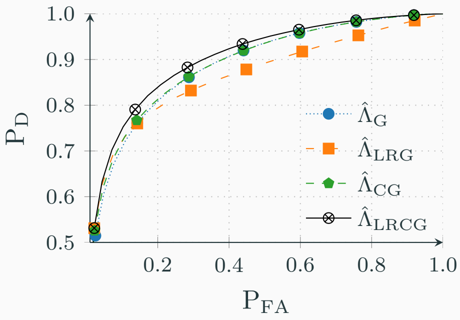
Results with a 5×5 sliding windows: Gaussian detectors



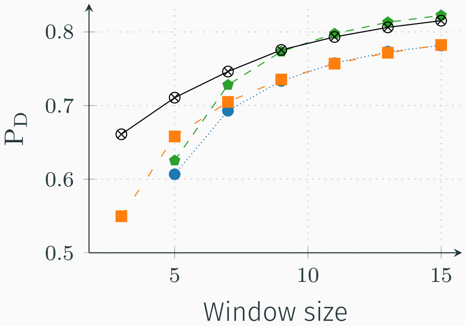
Results with a 5×5 sliding windows: Robust detectors



Performance curves ($p = 12$, $k = 3$)



ROC curves



P_D *vs window size at* $P_{FA} = 5\%$