Robust statistical framework for radar change detection applications - Part 2

Guillaume Ginolhac and Arnaud Breloy

and many thanks to: A. Mian, J-P. Ovarlez, A. Atto

IEEE Radar Conference - 26/09/2020

UNIVERSITÉ SAVOIE MONT BLANC

Université Paris Nanterre

Content

Content

Recap of Part 1

- **Change detection** in **multivariate time series**
- **Statistical framework**
  - Assume a distribution and meaningful feature parameters
  - Derive a decision function Λ
- **Gaussian** framework with **covariance matrix**
  - 2-step approaches: plug-in detectors (matrix distance) using the SCM
  - 1-step approaches: statistical criteria (CFAR property)
- **Part 2**: non-Gaussianity and dimensionality issues

Motivation of Part 2 (1/3): non-Gaussian data

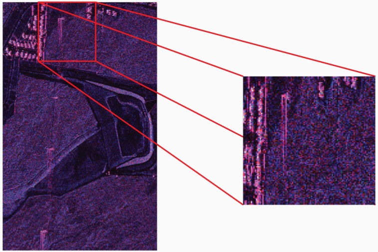
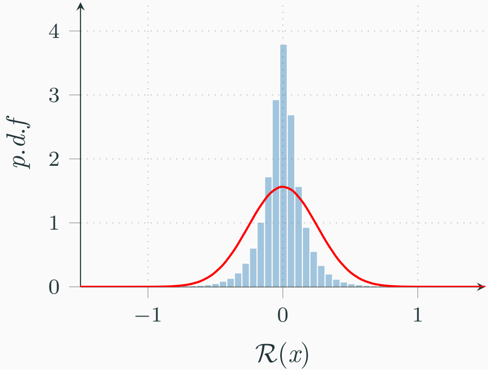


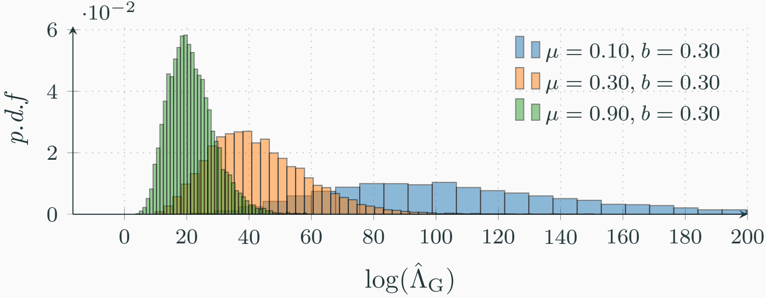
Figure 1: UAVSAR data (NASA/JPL-Caltech)



Gaussian models do not fit the empirical distribution of the data!

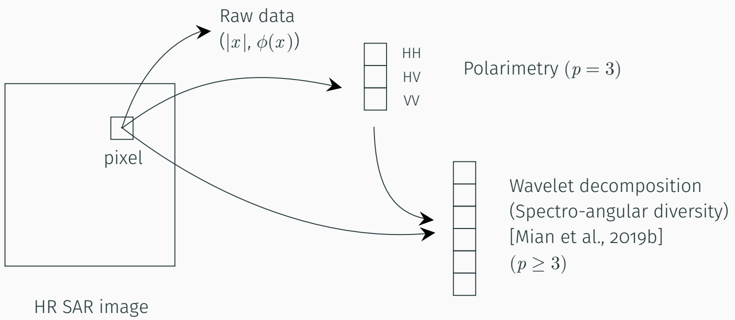
Motivation of Part 2 (2/3): issues in non-Gaussian context

$\mathbf{x}_i^t = \sqrt{\tau_i^t}\mathbf{z}_i^t$ where $\mathbf{z}_i^t \sim \mathbb{C}\mathcal{N}\left(\mathbf{0}_p, (0.5^{|i-j|})_{ij}\right)$ and $\tau_i^t \sim \Gamma(\mu, b)$, $p = 3$, $n = 10$, $T = 3$.



Gaussian detectors can perform poorly and lose properties (e.g. CFAR for GLRT)

Motivation of Part 2 (3/3): dimensionality issues?



- Improved CD performance with appropriate data transformation [Mian et al., 2017]
- Increasing p implies increasing n (patch dimension) $\Rightarrow$ lower CD resolution
- Can we achieve a reasonable trade-off?

PART 2

- **Sec.2**: Elliptical symmetric and compound Gaussian distributions
- **Sec.3**: M-estimators and robust plug-in detectors (2-step approach)
- **Sec.4**: Compound Gaussian GLRTs (1-step approach)
- **Sec.5-6**: Generalizations to structured covariance matrices

Content

Complex Elliptically Symmetric (CES) distributions

Definition **[Ollila et al., 2012]**

Let $\mathbf{x} \in \mathbb{C}^p$, $\mathbf{x}$ follows a CES ($\mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$) if its p.d.f. can be written

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

where

- $g : [0, \infty) \to [0, \infty)$ is the density generator
- $\boldsymbol{\mu}$ is the center of distribution
- $\boldsymbol{\Sigma}$ is the scatter matrix (full rank)

In general (finite second-order moment), $\mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^H\right] \propto \boldsymbol{\Sigma}$.

[Ollila et al., 2012] "Complex Elliptically Symmetric Distributions: Survey, New Results and Applications," IEEE Trans. on Signal Processing, vol. 60, no. 11, pp. 5597-5625, 2012.

CES characterizing property (1/3)

Stochastic representation theorem

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

$$\mathbf{x} \overset{d}{=} \boldsymbol{\mu} + \sqrt{\mathcal{Q}}\mathbf{A}\mathbf{u}$$

where

- $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^H$
- $\mathcal{Q} \geq 0$, is called the 2^{nd}-order modular variate:
  - independent of $\mathbf{u}$
  - whose p.d.f. only depends on g
- $\mathbf{u} \sim \mathcal{U}\left(\mathbb{C}\mathcal{S}^p\right)$, i.e., $\mathbf{u}$ is uniformly distributed on the unit sphere $\{\mathbf{x} \in \mathbb{C}^p \mid \|\mathbf{x}\| = 1\}$

CES characterizing property (2/3)

Properties

1. One-to-one relation between the p.d.f. of $\mathcal{Q}$ and g
2. Ambiguity: both $(\mathcal{Q}, \mathbf{A})$ and $(c^{-2}\mathcal{Q}, c\mathbf{A})$, $c > 0$ are stochastic representations of $\mathbf{x}$
   $\Rightarrow$ identifiability issues
3. Distribution of quadratic form:

$$Q(\mathbf{x}) \triangleq (\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \overset{d}{=} \mathcal{Q}$$

CES characterizing property (3/3)

Properties

4. Sample generation
   - draw $\mathcal{Q}$
   - draw $\mathbf{u}$ using $\mathbf{u} \overset{d}{=} \mathbf{g}/|\mathbf{g}|$, with $\mathbf{g} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{I})$
   - set $\mathbf{x} = \boldsymbol{\mu} + \sqrt{\mathcal{Q}}\mathbf{A}\mathbf{u}$

5. Practical interpretation
   - $\boldsymbol{\Sigma}$ accounts for correlations
   - $\mathcal{Q}$ accounts for amplitude fluctuations
   - models non-Gaussian distributions (e.g. heavy tails)

Fisher information matrix (FIM) for CES

Slepian-Bangs type formula **[Besson and Abramovich, 2013]**

Assuming

$$\mathbf{z} \sim \mathcal{CES}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

The FIM has for entries

$$\begin{aligned}
[\mathbf{F}(\boldsymbol{\theta})]_{ij} = \; & 2\gamma \mathfrak{Re}\left\{ \left.\frac{\partial \boldsymbol{\mu}^H(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left.\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j}\right|_{\boldsymbol{\theta}} \right\} \\
& + \alpha \mathrm{Tr}\left\{ \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left.\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left.\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_j}\right|_{\boldsymbol{\theta}} \right\} \\
& + \beta \mathrm{Tr}\left\{ \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left.\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta}} \right\} \mathrm{Tr}\left\{ \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left.\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_j}\right|_{\boldsymbol{\theta}} \right\}
\end{aligned}$$

where α, β and γ only depend on g.

Practical way to compute CRLBs

Compound Gaussian (CG) distributions (1/2)

Compound Gaussian (CG) distributions

An important subclass of CES, also called

- Spherically invariant random vectors (SIRVs) [Yao, 1973]
- Scale mixture of normal distributions [Andrews and Mallows, 1974]

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

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

where

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

Compound Gaussian (CG) distributions (2/2)

Comments

1. Indeed a subclass of the CES

$$\mathbf{x} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{\mathcal{Q}}\mathbf{A}\underbrace{\mathbf{g}/|\mathbf{g}|}_{\mathbf{u}} \stackrel{d}{=} \boldsymbol{\mu} + \underbrace{\sqrt{\mathcal{Q}}/|\mathbf{g}|}_{\sqrt{\tau}}\underbrace{\mathbf{A}\mathbf{g}}_{\mathbf{n}}$$

$\Rightarrow$ CG iff $\sqrt{\tau}$ and $\mathbf{n}$ are actually independent from this relation

2. Covariance matrix exists if $\mathbb{E}[\tau] < +\infty$, and

$$\mathbb{E}\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^H\right] = \mathbb{E}\left[\tau\right]\boldsymbol{\Sigma}$$

3. Identifiability: Both $(\sqrt{\tau}, \mathbf{n})$ and $(a\sqrt{\tau}, \mathbf{n}/a), \forall a > 0$ leads to same CG dist. for $\mathbf{x}$

Examples (1/2)

t-distribution

- CG representation with $\tau^{-1} \sim \Gamma(\nu/2, 2/\nu)$, where $\nu > 0$ (d.o.f.)
- $\nu = 1 \implies$ complex Cauchy dist.
- $\nu \to \infty \implies$ CN dist.

K-distribution

- CG representation with $\tau \sim \Gamma(\nu, 1/\nu)$, where $\nu > 0$
- $\nu \downarrow \implies$ heavy-tailed dist
- $\nu \to \infty \implies$ CN dist

Examples (2/2)

Generalized Gaussian distribution

- CES representation with $\mathcal{Q} \stackrel{d}{=} G^{1/s}$ where $G \sim \Gamma(m/s, \eta)$, $s, \eta > 0$
- $s = 1 \implies$ CN dist.
- Heavier tails for $s < 1$ and lighter tails for $s > 1$

CG with deterministic textures

- CG representation with unknown deterministic textures $\{\tau_i\}$
- Conditional representation

$$\mathbf{x}_i | \tau_i \sim \mathbb{C}\mathcal{N}(\boldsymbol{\mu}, \tau_i \boldsymbol{\Sigma})$$

- Practical to derive robust processes with unknown f_τ (or g)

Practical use of CES/CG distributions

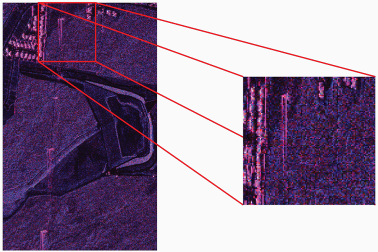
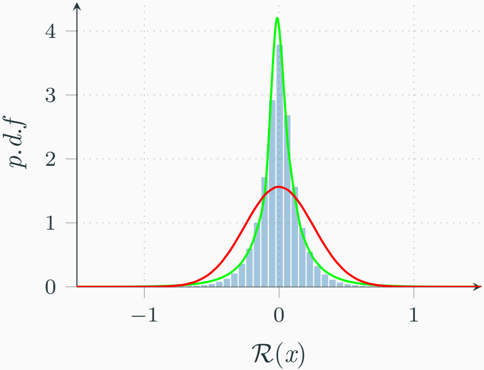


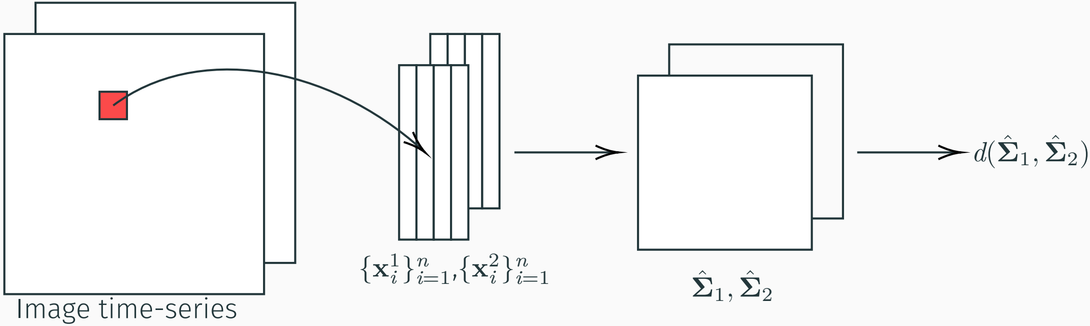
Figure 2: UAVSAR data (NASA/JPL-Caltech)



Good fit to the empirical distribution of the data!

Content

2-step CD: recap



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

2-step CD: from Gaussian to CES

Gaussian plug-in detectors

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

Issues when samples are CES

- The SCM is an inaccurate estimator
- The CES density generator g is unknown in practice
- We need robust estimators of the covariance matrix

Recall on the sample covariance matrix

Gaussian model

$\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ has the p.d.f.:

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

Sample covariance matrix (SCM)

Maximum likelihood estimator of the covariance matrix:

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

$\rightarrow$ Not robust to heavy tails (nor outliers).

Covariance matrix estimation in CES (1/2)

Maximum likelihood estimator (known g)

Let $\{\mathbf{x}_i\}_{i=1}^n \in (\mathbb{C}^p)^n$ be a n-sample following $\mathbf{x} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g)$

- MLE: $\widehat{\boldsymbol{\Sigma}}_{\text{MLE}}$ that minimizes the negative log-likelihood function

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

- Solution to the fixed-point equation

$$\widehat{\boldsymbol{\Sigma}}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} \varphi(\mathbf{x}_i^H \widehat{\boldsymbol{\Sigma}}_{\text{MLE}}^{-1} \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^H$$

with $\varphi(t) = -g'(t)/g(t)$.

Covariance matrix estimation in CES (2/2)

M-Estimators (unknown g)

(*PDF not specified* $\Rightarrow$ *M-estimators can be used instead of MLE*)

A complex M-estimator of $\mathbf{\Sigma}$ is defined as the solution of

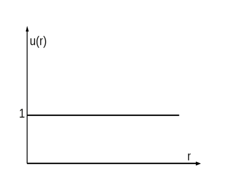
$$\widehat{\mathbf{\Sigma}}_M = \frac{1}{n}\sum_{i=1}^{n} u\left(\mathbf{x}_i^H \widehat{\mathbf{\Sigma}}_M^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H,$$

for a given weight function u (not necessarily linked to g, φ).

Examples of M-estimators (1/2)

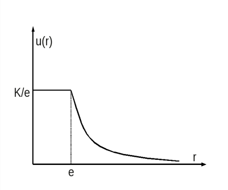
SCM

$$u(r) = 1$$



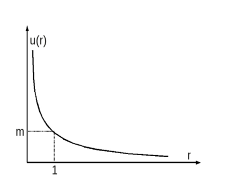
Huber

$$u(r) = \begin{cases} A/e \text{ if } r <= e \\ A/r \text{ if } r > e \end{cases}$$



Tyler

$$u(r) = \frac{p}{r}$$



Remarks:

- Huber = mix between SCM and Tyler
- Performance/robustness trade-off

Tyler Estimator:

$$\mathbf{V} = \frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^H}{\mathbf{x}_i^H\mathbf{V}^{-1}\mathbf{x}_i}$$

Examples of M-estimators (2/2)

Other option: assume a specific g (not necessarily true)

Example : t-distribution with degree of freedom d

$$g(t) = \left(1 + \frac{2\,t}{d}\right)^{-(2p+d)/2}$$

Derive an M-estimator as its corresponding MLE:

$$\widehat{\boldsymbol{\Sigma}}_d = \frac{d+p}{n} \sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^H}{d + \mathbf{x}_i^H \widehat{\boldsymbol{\Sigma}}_d^{-1} \mathbf{x}_i}$$

$\Rightarrow$ trade-off between SCM ($d \longrightarrow \infty$) and Tyler ($d \longrightarrow 0$)

M-estimators: some key properties (1/2)

1. Computation with fixed-point algorithm (EM and MM interpretation)

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

2. Existence, uniqueness, algorithm convergence subject to conditions on u and $n > p$
   - Real case [Maronna, 1976, Tyler, 1987]
   - Complex case [Pascal et al., 2008, Ollila et al., 2012]
   - Using g-convexity [Wiesel, 2012]
3. Several asymptotic characterization
   - Standard Gaussian asymptotic [Ollila et al., 2012]
   - Asymptotic Wishart equivalent [Drašković and Pascal, 2018]
   - Large (n, p) regime [Zhang et al., 2014, Couillet et al., 2015]
   - PAC bounds [Soloveychik and Wiesel, 2015]
   - Comparison with Cramer-Rao bounds [Greco and Gini, 2013, Breloy et al., 2019]

M-estimators: some key properties (2/2)

4. Asymptotically unbiased and consistent estimators up to a scaling

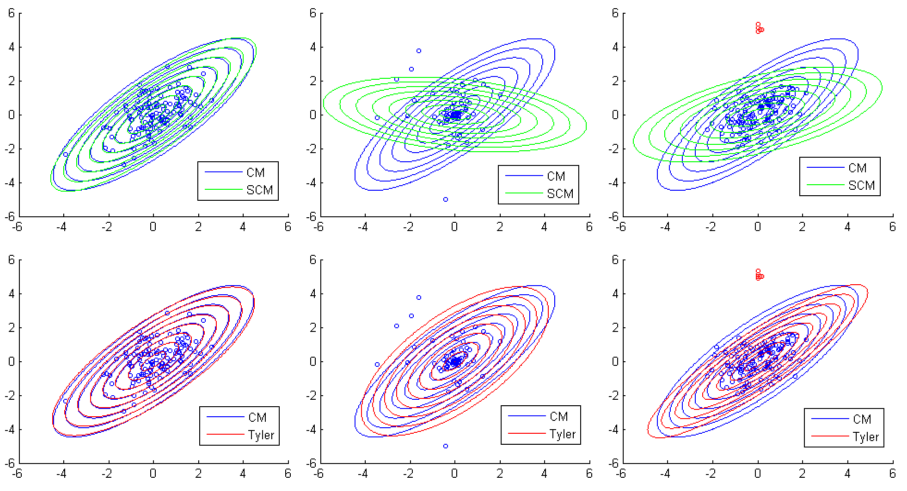
   $\Rightarrow$ Estimators of the *shape matrix*

5. Robust over the class of CES (Tyler is even "distribution-free")

6. Robust to outliers [Maronna, 1976]
   - Influence function
   - Breakdown point

7. If the data is Gaussian: little loss compared to the SCM

Interest: visual examples with SCM and Tyler



Alternate way to demonstrate uniqueness: geodesic convexity

g-convexity

- Also referred to as super-convexity or arcwise connectivity
- Extends the concept of convexity to geodesic curves (link with Riemannian geometry)
- Used in many recent references about covariance estimation [Wiesel, 2011, Wiesel, 2012, Zhang et al., 2013, Ollila and Tyler, 2014, Duembgen and Tyler, 2016, Mian et al., 2019a, Breloy et al., 2019]

Geodesic curve on $\mathcal{H}_p^{++}$

Let the pair $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2 \in \mathcal{H}_p^{++}$, define the curve

$$\boldsymbol{\Sigma}(t) = \boldsymbol{\Sigma}_1^{1/2}(\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1/2})^t\boldsymbol{\Sigma}_1^{1/2}$$

(shortest path between $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ on $\mathcal{H}_p^{++}$ endowed with its natural metric)

g-convexity on $\mathcal{H}_p^{++}$: definitions

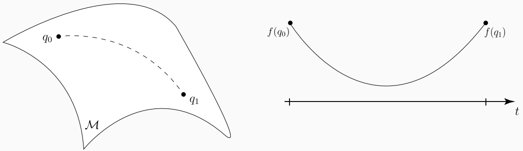
g-convex set of $\mathcal{H}_p^{++}$

A set $\mathcal{S} \in \mathcal{H}_p^{++}$ is g-convex if for any $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2 \in \mathcal{S}$, $\mathbf{\Sigma}(t) \in \mathcal{S}\ \forall t$

g-convex function

Let $\mathcal{S} \in \mathcal{H}_p^{++}$ be a g-convex set, a function f is g-convex on $\mathcal{S}$ if for any pair $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2 \in \mathcal{S}$

$$f(\mathbf{\Sigma}(t)) \leq t f(\mathbf{\Sigma}_1) + (1-t) f(\mathbf{\Sigma}_2),\ \forall t \in [0,1]$$



g-convexity: key properties

Propositions

- If f is geodesically convex on $\mathcal{H}_p^{++}$, any local minimum is a global minimum.
- If a minimum is obtained in $\mathcal{H}_p^{++}$ then the set of all minimums form a g-convex subset of $\mathcal{H}_p^{++}$.
- If f is strictly g-convex and a minimum is obtained on $\mathcal{H}_p^{++}$, then it is a unique minimum.

g-convexity of M-estimators cost functions

Proposition

Let $\rho(t)$ be a non decreasing continuous function such that

$$r(x) = \rho(e^x)$$

is convex, then

$$\mathcal{L}(\mathbf{\Sigma}) = \frac{1}{n}\sum_{i=1}^{n}\rho(\mathbf{x}^H\mathbf{\Sigma}^{-1}\mathbf{x}) + \ln|\mathbf{\Sigma}|$$

is g-convex on $\mathcal{H}_p^{++}$.

Examples:

- CES log-likelihoods
- Costs linked to M-estimations: $\rho(t) = \ln|t|$ for Tyler estimator

Conclusions

On CES/CG models

- Flexible family of multivariate distributions
- Good empirical fit to many datasets
- M-estimators: robust estimators suited to this family

On g-convexity on $\mathcal{H}_p^{++}$

- Useful to derive meaningful costs/penalties for robust covariance matrix estimation
- Provides guarantees on existence/uniqueness of the solutions

Can we now apply M-estimators to plug-in change detectors ?

Recap on UAVSAR data

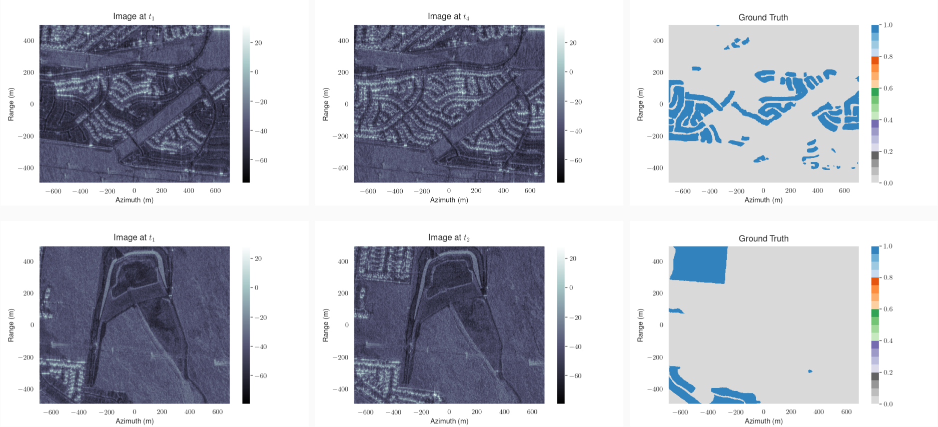


Figure 3: UAVSAR SanAnd_26524_03 scenes 1 and 2 [Ratha et al., 2017, Nascimento et al., 2019]

Compared detectors

- $\hat{\Lambda}_{\mathrm{G}}$: Gaussian GLRT as baseline
- Plug-in Rao distance

$$d_{\mathrm{Rao}}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \alpha \sum_{i=1}^{p} \log^2 \lambda_i + \beta \left(\sum_{i=1}^{p} \log \lambda_i \right)^2 \text{ with } \{\lambda_i\}_{i=1}^{p} = \mathrm{eig}(\mathbf{\Sigma}_1^{-1} \mathbf{\Sigma}_2)$$

  - $\hat{\Lambda}_{\mathcal{RG}}$: SCMs, $\alpha = 1$, $\beta = 0$
  - $\hat{\Lambda}_{\mathcal{RE}}$: MLE of t-distribution (d.o.f. $d = 3$), $\alpha = \frac{d+p}{d+p+1}$, $\beta = \alpha - 1$
- Plug-in Wasserstein distance

$$d_{\mathrm{W}}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \mathrm{Tr}\left\{ \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2 - 2 \left(\mathbf{\Sigma}_2^{1/2} \mathbf{\Sigma}_1 \mathbf{\Sigma}_2^{1/2} \right)^{1/2} \right\}$$

  - $\hat{\Lambda}_{\mathcal{WG}}$: computed with SCMs
  - $\hat{\Lambda}_{\mathcal{WE}}$: computed with Tyler M-estimator (scale averaged)

Results

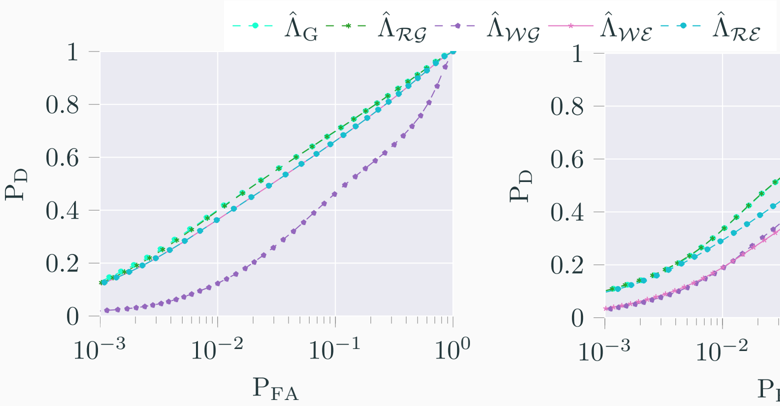


Figure 4: ROC plots using a 5×5 local window for the scenes 1 and 2.

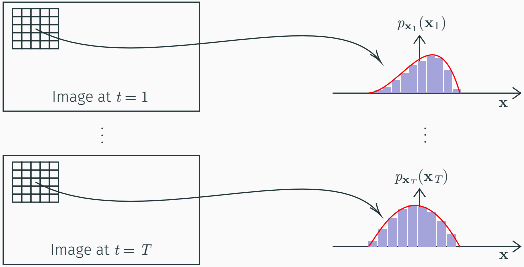
Conclusion on the 2-step approach for CES data

- **M-estimators**: are very good at robustly estimating $\boldsymbol{\Sigma}$
- **2-step CD application**: not always beneficial
  - Robustness: not sensitive to a change in few pixels in the patch (edges)
  - 2-step CD: does not grasp modular-variate/textures variations
  - M-estimators even mitigate their impact
    e.g., Tyler makes a change in scale vanish, while we want to detect it

We need to turn to statistical criteria to fully leverage CES models !

Content

Brief recap on 1-step statistical detection



- Set a probabilistic model

$$\mathbf{x}_i^t \sim p_{\mathbf{x}_i^t}(\mathbf{x}_i^t; \boldsymbol{\theta}_t; \boldsymbol{\Phi}_t),$$

$\{\boldsymbol{\theta}_t, \boldsymbol{\Phi}_t\}$: interest/nuisance parameters

- Detect change in $\boldsymbol{\theta}_t \Leftrightarrow$ binary hypothesis test

$$\begin{cases} \mathrm{H}_0 : & \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t'} \quad \& \quad \boldsymbol{\Phi}_t \neq \boldsymbol{\Phi}_{t'}, \quad \forall (t, t') \\ \mathrm{H}_1 : & \boldsymbol{\theta}_t \neq \boldsymbol{\theta}_{t'} \quad \& \quad \boldsymbol{\Phi}_t \neq \boldsymbol{\Phi}_{t'}, \quad \forall (t, t') \end{cases}$$

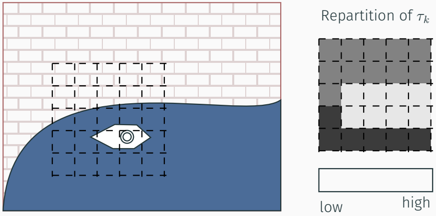
- Derive a criterion (GLRT, Rao, Wald, etc.)

Model for CD within CG

Model

CES with no assumption on density $g \longrightarrow$ CG with unknown deterministic textures:

$$\mathbf{x}_i^t \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \tau_i^t \boldsymbol{\Sigma}_t)$$



Intuition: $\boldsymbol{\Sigma}_t$ captures local correlations and τ_i^t captures power fluctuations

Derivation of the GLRT

- **The MLE is Tyler's M-estimator**, why not plug it in in the formula of Λ_G ?
  - Not the true GLRT: 2-step approach using GLRT as a distance
  - Tyler's scale ambiguity $\rightarrow$ losing power variations
  - τ is a new parameter that can also be tested individually
- 3 detection tests [Mian et al., 2019a]

$$\begin{cases} \text{Covariance only:} & \boldsymbol{\theta}_t = \{\boldsymbol{\Sigma}_t\} & \& & \boldsymbol{\Phi}_t = \{\boldsymbol{\tau}_t\} \\ \text{Covariance and textures:} & \boldsymbol{\theta}_t = \{\boldsymbol{\tau}_t, \boldsymbol{\Sigma}_t\} & \& & \boldsymbol{\Phi}_t = \varnothing \\ \text{Textures only:} & \boldsymbol{\theta}_t = \{\boldsymbol{\tau}_t\} & \& & \boldsymbol{\Phi}_t = \{\boldsymbol{\Sigma}_t\} \end{cases}$$

GLRT's expressions

Covariance only ($\theta_t = \{\mathbf{\Sigma}_t\}$ & $\mathbf{\Phi}_t = \{\boldsymbol{\tau}_t\}$)

$$\hat{\Lambda}_{\mathbb{C}\mathcal{AE}} = \frac{\left|\hat{\mathbf{\Sigma}}_0^{\mathrm{TE}}\right|^{Tn}}{\prod_{t=1}^{T}\left|\hat{\mathbf{\Sigma}}_t^{\mathrm{TE}}\right|^{n}} \prod_{i=1}^{n}\prod_{t=1}^{T} \frac{\left(q(\hat{\mathbf{\Sigma}}_0^{\mathrm{TE}}, \mathbf{x}_i^t)\right)^p}{\left(q(\hat{\mathbf{\Sigma}}_t^{\mathrm{TE}}, \mathbf{x}_i^t)\right)^p} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda,$$

where: $\hat{\mathbf{\Sigma}}_t^{\mathrm{TE}} = f_t^{\mathrm{TE}}(\hat{\mathbf{\Sigma}}_t^{\mathrm{TE}})$, $\hat{\mathbf{\Sigma}}_0^{\mathrm{TE}} = \frac{1}{T}\sum_{t=1}^{T} f_t^{\mathrm{TE}}(\hat{\mathbf{\Sigma}}_0^{\mathrm{TE}})$, $q(\mathbf{\Sigma}, \mathbf{x}) = \mathbf{x}^{\mathrm{H}}\mathbf{\Sigma}^{-1}\mathbf{x}$ and

$$f_t^{\mathrm{TE}}(\mathbf{\Sigma}) = \frac{p}{n}\sum_{i=1}^{n} \frac{\mathbf{x}_i^t \mathbf{x}_i^{t\,\mathrm{H}}}{q(\mathbf{\Sigma}, \mathbf{x}_i^t)}.$$

Property of global maximum thanks to g-convexity [Wiesel, 2012]

GLRT's expressions

Covariance and texture ($\theta_t = \{\boldsymbol{\tau}_t, \boldsymbol{\Sigma}_t\}$ & $\boldsymbol{\Phi}_t = \varnothing$)

$$\hat{\Lambda}_{\mathrm{MT}} = \frac{\left|\hat{\boldsymbol{\Sigma}}_0^{\mathrm{MT}}\right|^{Tn}}{\prod_{t=1}^{T}\left|\hat{\boldsymbol{\Sigma}}_t^{\mathrm{TE}}\right|^{n}} \prod_{i=1}^{n} \frac{\left(\sum_{t=1}^{T} q\left(\hat{\boldsymbol{\Sigma}}_0^{\mathrm{MT}}, \mathbf{x}_i^t\right)\right)^{Tp}}{T^{Tp}\prod_{t=1}^{T}\left(q\left(\hat{\boldsymbol{\Sigma}}_t^{\mathrm{TE}}, \mathbf{x}_i^t\right)\right)^{p}} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda \,,$$

where:

$$\hat{\boldsymbol{\Sigma}}_0^{\mathrm{MT}} = f_{n,T}^{\mathrm{MT}}\left(\hat{\boldsymbol{\Sigma}}_0^{\mathrm{MT}}\right) = \frac{p}{n}\sum_{i=1}^{n} \frac{\sum_{t=1}^{T} \mathbf{x}_i^t (\mathbf{x}_i^t)^{\mathrm{H}}}{\sum_{t=1}^{T} q\left(\hat{\boldsymbol{\Sigma}}_0^{\mathrm{MT}}, \mathbf{x}_i^t\right)} \,.$$

Property of global maximum thanks to g-convexity [Wiesel, 2012]

GLRT's expressions

Texture only ($\theta_t = \{\boldsymbol{\tau}_t\}$ & $\boldsymbol{\Phi}_t = \boldsymbol{\Sigma}_t$)

$$\hat{\Lambda}_{\text{Tex}} = \prod_{t=1}^{T} \frac{\left|\hat{\boldsymbol{\Sigma}}_t^{\text{Tex}}\right|^n}{\left|\hat{\boldsymbol{\Sigma}}_t^{\text{TE}}\right|^n} \prod_{i=1}^{n} \frac{\left(\sum_{t=1}^{T} q\left(\hat{\boldsymbol{\Sigma}}_t^{\text{Tex}}, \mathbf{x}_i^t\right)\right)^{Tp}}{T^{Tp} \prod_{t=1}^{T} \left(q\left(\hat{\boldsymbol{\Sigma}}_t^{\text{TE}}, \mathbf{x}_i^t\right)\right)^p} \underset{\text{H}_0}{\overset{\text{H}_1}{\gtrless}} \lambda \,,$$

where:

$$\hat{\boldsymbol{\Sigma}}_t^{\text{Tex}} = f_{n,T,t}^{\text{Tex}}\left(\hat{\boldsymbol{\Sigma}}_1^{\text{Tex}}, \dots, \hat{\boldsymbol{\Sigma}}_T^{\text{Tex}}\right) = \frac{Tp}{n} \sum_{i=1}^{n} \frac{\mathbf{x}_i^t (\mathbf{x}_i^t)^{\text{H}}}{\sum_{t'=1}^{T} q\left(\hat{\boldsymbol{\Sigma}}_{t'}^{\text{Tex}}, \mathbf{x}_i^t\right)} \,.$$

Property of global maximum thanks to g-convexity [Wiesel, 2012]

CFAR properties

CFARness w.r.t. shape matrix

$\hat{\Lambda}_{\mathrm{MT}}$ and $\hat{\Lambda}_{\mathbb{C}\mathcal{A}\mathcal{E}}$ are CFAR matrix while $\hat{\Lambda}_{\mathrm{Tex}}$ is not.

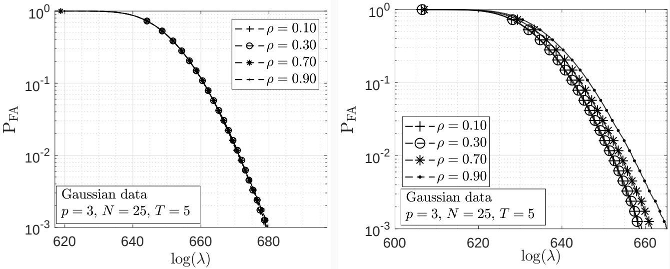


Figure 5: PFA versus threshold with $\mathbf{x}_i^t \sim \mathbb{C}\mathcal{N}\left(\mathbf{0}_p, (\rho^{|i-j|})_{ij}\right)$.

CFAR properties

CFARness w.r.t. texture parameters

$\hat{\Lambda}_{\mathrm{MT}}$, $\hat{\Lambda}_{\mathbb{C}\mathcal{AE}}$ and $\hat{\Lambda}_{\mathrm{Tex}}$ are CFAR texture.

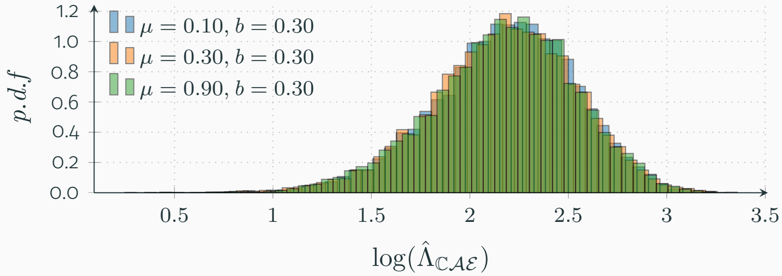


Figure 5: Test statistic histogram with $\mathbf{x}_i^t = \sqrt{\tau_i^t}\mathbf{z}_i^t$ where $\mathbf{z}_i^t \sim \mathbb{C}\mathcal{N}\left(\mathbf{0}_p, (0.3^{|i-j|})_{ij}\right)$ and $\forall(i,t)$, $\tau_i^t = \tau_i \sim \Gamma(\nu, b)$

Results on UAVSAR data

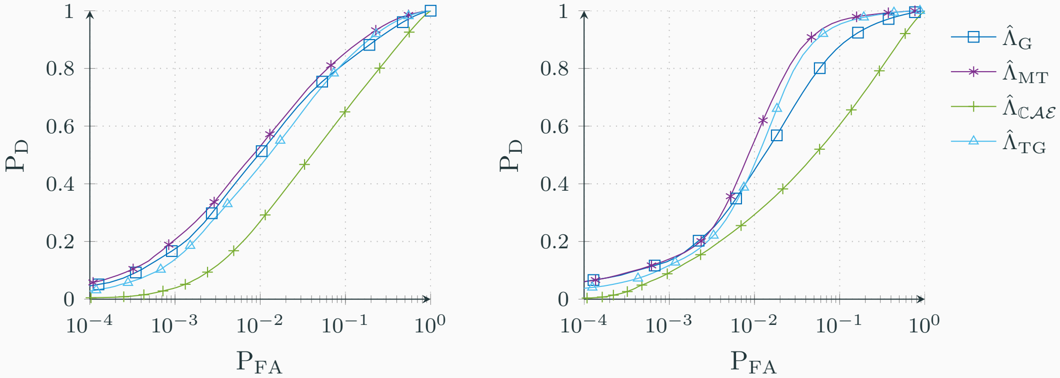


Figure 6: ROC plots using a 5×5 local window for the scenes 1 and 2.

Some concluding remarks

- **2-steps approaches**: do not fully leverage CES models
- **GLRT for covariance/texture testing under CG modelling**
  - Improved performance on SAR data
  - CFAR properties
- **Extensions**
  - CES case can be transposed but is more specific (assumed g)
  - t_1, Wald, Rao, etc., remain to be derived

Remaining issue

- Dealing with high p with reasonable n
- Regularization or structured covariance matrices

Content

Motivation through CD in SAR ITS

Dimensionality issue

- Wavelet transformations highlight interesting phenomenons but increase p
- $n > p$ required for M-estimators to exist
- $n \sim 2p$ is a common rule of thumb to expect a correct estimation

$\Rightarrow$ Decreases the spatial resolution in CD

A possible solution

- Introduce prior information on the covariance structure
- Reduce the estimation problem dimension: better performances with low n

Examples of covariance matrices structures (1/2)

- **Linear and convex sets** [Meriaux et al., 2019, Soloveychik and Wiesel, 2014]

$$\boldsymbol{\Sigma} = \sum_{i=1}^{L} \alpha_i \mathbf{B}_i \text{ with } \forall i,\ \alpha_i \in \mathbb{R}$$

Common in radar signal processing (Toeplitz, sum-of-rank-1, blocks...)

- **Group symmetric structure** [Soloveychik et al., 2015]

$$\boldsymbol{\Sigma} = \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^H \text{ for } \mathbf{H} \in \mathcal{G}$$

Induced by symmetries of the sensing system

Examples of covariance mtrices structures (2/2)

- **Kronecker products** [Wiesel, 2012, Breloy et al., 2016]

$$\boldsymbol{\Sigma} = \mathbf{A} \otimes \mathbf{B}$$

Common in MIMO systems

- **Factor models** (spiked, low-rank) [Sun et al., 2015]

$$\boldsymbol{\Sigma} = \sum_{r=1}^{k} \lambda_r \mathbf{v}_r \mathbf{v}_r^H + \sigma^2 \mathbf{I}$$

Ubiquitous in radar, finance, bioinformatics, ...

Robust structured estimation: EXIP approaches [Meriaux et al., 2019]

- **Parameterization** of the structure $\boldsymbol{\Sigma} = \boldsymbol{\mathcal{R}}(\boldsymbol{\theta})$

- **2-step estimation** procedure
  1. Compute an M-estimator $\widehat{\boldsymbol{\Sigma}}_M$
  2. Refine the estimate with the projection $\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} \mathcal{J}_{\hat{\boldsymbol{\Sigma}}_m, \hat{\boldsymbol{\Sigma}}}(\boldsymbol{\theta})$, with

$$\mathcal{J}_{\hat{\boldsymbol{\Sigma}}_m, \hat{\boldsymbol{\Sigma}}}(\boldsymbol{\theta}) = \alpha \text{Tr}\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\Sigma}}_m - \boldsymbol{\mathcal{R}}(\boldsymbol{\theta})\right)\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\Sigma}}_m - \boldsymbol{\mathcal{R}}(\boldsymbol{\theta})\right)\right) + \beta\left[\text{Tr}\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\Sigma}}_m - \boldsymbol{\mathcal{R}}(\boldsymbol{\theta})\right)\right)\right]^2$$

  $\hat{\boldsymbol{\Sigma}}$ is a consistent estimator, (α, β) build the FIM [Besson and Abramovich, 2013]

- **Performance**
  - Theoretically guarantees, asymptotic (m)-efficiency
  - Sub-optimal at low sample support

Robust structured estimation: optimization approaches

$$\begin{aligned} \underset{\boldsymbol{\Sigma}}{\text{minimize}} \quad & \mathcal{L}(\boldsymbol{\Sigma}) \\ \text{subject to} \quad & \boldsymbol{\Sigma} \in \mathcal{S} \end{aligned}$$

Convexity, g-convexity

- recasting meaningful problems [Soloveychik and Wiesel, 2014]
- guarantees on global optimality for some structures [Wiesel and Zhang, 2015]

Majorization-minimization (MM) [Sun et al., 2016a]

- Closed-form iteration, scalable algorithms
- Most structures have been tackled [Sun et al., 2016b, Breloy et al., 2016]

SAR data: a focus on low-rank models

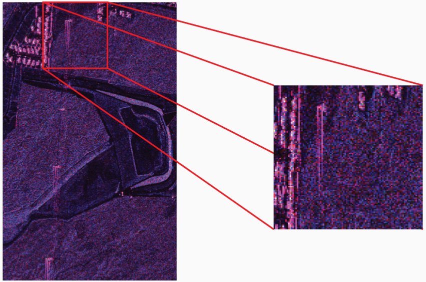


Figure 7: UAVSAR data (NASA/JPL-Caltech)

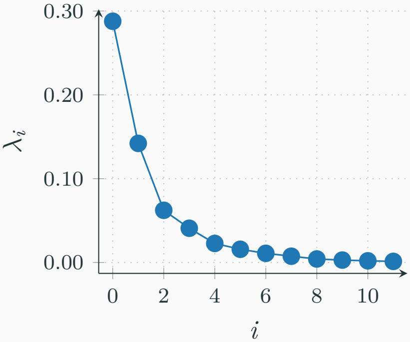


Figure 8: Spectrum ($p = 12$ wavelets)

Extending Tyler's estimator to low-rank models

$$\begin{aligned} \underset{\boldsymbol{\Sigma}}{\text{minimize}} \quad & \frac{p}{n} \sum_{i=1}^{n} \ln \left(\mathbf{x}_i^H \boldsymbol{\Sigma}^{-1} \mathbf{x}_i \right) + \ln |\boldsymbol{\Sigma}| \\ \text{subject to} \quad & \boldsymbol{\Sigma} = \mathbf{V}\mathbf{D}\mathbf{V}^H + \sigma^2 \mathbf{I}_p \\ & \mathbf{V}^H \mathbf{V} = \mathbf{I}_k \end{aligned}$$

Issues

- No explicit solution
- Non-convex (nor g-convex)
- Requires iterative algorithms to evaluate local maximum $\rightarrow$ MM algorithm

The MM Algorithm (1/3) [Hunter and Lange, 2004, Sun et al., 2016a]

- Consider the optimization problem

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X}, \end{aligned}$$

where f is too complex to handle directly

- The idea is to successively minimize an approximation $g(\mathbf{x}|\mathbf{x}_t)$

$$\mathbf{x}_{t+1} = \underset{\mathbf{x} \in \mathcal{X}}{\text{argmin}} \; g(\mathbf{x}|\mathbf{x}_t)$$

hoping the sequence $\{\mathbf{x}_t\}$ will converge to an optimal point of f

- The MM algorithm provides
  - The guidelines for the construction of such function g
  - The conditions to ensure the success of this method

The MM Algorithm (2/3) [Hunter and Lange, 2004, Sun et al., 2016a]

Construction rules for the surrogate function g

(A1) Equality at the considered point

$$g(\mathbf{y}|\mathbf{y}) = f(\mathbf{y}) \;\; \forall \, \mathbf{y} \in \mathcal{X}$$

(A2) "Majorization"

$$f(\mathbf{x}) \leq g(\mathbf{x}|\mathbf{y}) \;\; \forall \, \mathbf{x}, \mathbf{y} \in \mathcal{X}$$

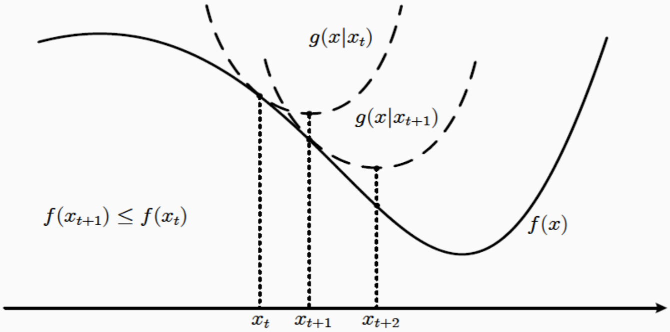
(A3) Equality of directional derivatives

$$g'(\mathbf{x}, \mathbf{y}; \mathbf{d})|_{\mathbf{x}=\mathbf{y}} = f'(\mathbf{y}; \mathbf{d}) \;\; \forall \, \mathbf{d} \text{ with } \mathbf{y} + \mathbf{d} \in \mathcal{X}$$

(A4) $g(\mathbf{x}|\mathbf{y})$ is continuous in $\mathbf{x}$ and in $\mathbf{y}$

The MM Algorithm (3/3) [Hunter and Lange, 2004, Sun et al., 2016a]

"Iteratively minmizing a smooth local tight upperbound of the objective"



$$\mathbf{x}_{t+1} = \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} \; g(\mathbf{x}|\mathbf{x}_t)$$

Applying MM to our problem (1/2)

Majorization: "Gaussian" surrogate

$$\frac{p}{n}\sum_{i=1}^{n}\ln\left(\mathbf{x}_i^H\boldsymbol{\Sigma}^{-1}\mathbf{x}_i\right)+\ln|\boldsymbol{\Sigma}| \;\leq\; \frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i^H\boldsymbol{\Sigma}^{-1}\mathbf{x}_i}{\mathbf{x}_i^H\boldsymbol{\Sigma}_t^{-1}\mathbf{x}_i}+\ln|\boldsymbol{\Sigma}|+\text{const.}$$

Similar to a Gaussian likelihood with reweighted samples (IRLS)

Minimization: MLE for Gaussian factor models [Tipping and Bishop, 1999]

Let $\{\mathbf{x}_i\}_{i=1}^{n}$ follow $\mathbf{x}\sim\mathbb{C}\mathcal{N}(\mathbf{0},\mathbf{V}\mathbf{D}\mathbf{V}^H+\sigma^2\mathbf{I}_p)$, the MLE of the covariance matrix is

$$\widehat{\boldsymbol{\Sigma}} \;=\; \mathcal{T}_k\left\{\widehat{\boldsymbol{\Sigma}}_{\text{SCM}}\right\} \;\overset{\text{EVD}}{=}\; \widehat{\mathbf{V}}\text{diag}\left(\left[d_1,\ldots,d_k,\hat{\sigma}^2,\ldots,\hat{\sigma}^2\right]\right)\widehat{\mathbf{V}}^H$$

with $\widehat{\boldsymbol{\Sigma}}_{\text{SCM}} \overset{\text{EVD}}{=} \widehat{\mathbf{V}}\text{diag}\left([d_1,\ldots,d_p]\right)\widehat{\mathbf{V}}^H$ and $\hat{\sigma}^2=\text{mean}\left([d_{k+1},\ldots,d_p]\right)$

Applying MM to our problem (2/2)

Algorithm 1 MM for low-rank structured Tyler's estimator

repeat

Compute the SCM of normalized samples $\mathbf{x}_i/(\sqrt{\mathbf{x}_i^H \boldsymbol{\Sigma}_t^{-1} \mathbf{x}_i / p})$

Update $\boldsymbol{\Sigma}_{t+1}$ by averaging $(p-k)$ last eigenvalues

until convergence

Alternatives

- EM-type/BCD w. textures $((\{\tau_i\}, \boldsymbol{\Sigma}) \rightarrow$ yields the same iterations !
- Riemannian optimization (beneficial at low sample support) [Bouchard et al., 2020a]

Content

Application to GLRT for CD

GLRT

Given the data $\{\{\mathbf{x}_i^t\}_{i=1}^n\}_{t=1}^T$, model $p_{\mathbf{x}}$ and test parameters $\{\boldsymbol{\theta}, \boldsymbol{\Phi}\}$, the GLRT is

$$\hat{\Lambda}_{\text{GLRT}} = \frac{\max\limits_{\{\boldsymbol{\theta}_t, \boldsymbol{\Phi}_t\}_{t=1}^T} p_{\mathbf{x}}\left(\{\{\mathbf{x}_i^t\}_{i=1}^n\}_{t=1}^T \; ; \; \{\boldsymbol{\theta}_t, \boldsymbol{\Phi}_t\}_{t=1}^T \mid \mathrm{H}_1\right)}{\max\limits_{\boldsymbol{\theta}_0, \{\boldsymbol{\Phi}_t\}_{t=1}^T} p_{\mathbf{x}}\left(\{\{\mathbf{x}_i^t\}_{i=1}^n\}_{t=1}^T \; ; \; \boldsymbol{\theta}_0, \{\boldsymbol{\Phi}_t\}_{t=1}^T \mid \mathrm{H}_0\right)} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda.$$

For CD with low-rank models

- No closed form expression of the MLE
- Optimization to evaluate the maximum over both hypotheses

Low-rank Gaussian CD (1/2) [Ben Abdallah et al., 2019]

Model and parameters

$$\mathbf{x}_i^t \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}_k^t + \sigma_t^2\, \mathbf{I}\right)$$

where

$$\mathcal{L}_{\mathcal{G}}\left(\{\mathbf{x}_i\}_{i=1}^n | \mathbf{\Sigma}\right) \propto \prod_{i=1}^n |\mathbf{\Sigma}|^{-1} \exp\left(-\mathbf{x}_i^H \mathbf{\Sigma}^{-1} \mathbf{x}_i\right)$$

Hypothesis test

$$\begin{aligned} \mathrm{H}_0 : \quad & \boldsymbol{\theta}_0 && = \left\{\mathbf{\Sigma}_k^0, \sigma_0^2\right\} \\ \mathrm{H}_1 : \quad & \{\boldsymbol{\theta}_t\}_{t=1}^T && = \left\{\mathbf{\Sigma}_k^t, \sigma_t^2\right\}_{t=1}^T \end{aligned}$$

Low-rank Gaussian CD (2/2) [Ben Abdallah et al., 2019]

Expression of the GLRT: no closed form

$$\hat{\Lambda}_{\mathrm{LRG}} = \frac{\mathcal{L}\left(\{\{\mathbf{x}_i^t\}_{i=1}^n\}_{t=1}^T \,|\mathrm{H}_1; \mathcal{T}_k\{\hat{\mathbf{\Sigma}}_1\}, \ldots, \mathcal{T}_k\{\hat{\mathbf{\Sigma}}_T\}\right)}{\mathcal{L}\left(\{\{\mathbf{x}_i^t\}_{i=1}^n\}_{t=1}^T \,|\mathrm{H}_0; \mathcal{T}_k\{\hat{\mathbf{\Sigma}}_0\}\right)} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda.$$

where $\hat{\mathbf{\Sigma}}_t = \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i^t(\mathbf{x}_i^t)^H$. From $\mathbf{\Sigma} \overset{\mathrm{EVD}}{=} \mathbf{U}\mathrm{diag}(\mathbf{d})\mathbf{U}^H$, we obtain $\mathcal{T}_k\{\mathbf{\Sigma}\}$:

$$\mathcal{T}_k\{\mathbf{\Sigma}\} = \mathbf{U}\mathrm{diag}(\tilde{\mathbf{d}})\mathbf{U}^H$$

where $\tilde{\mathbf{d}} = [d_1, \ldots, d_k, \hat{\sigma}_t^2, \ldots, \hat{\sigma}_t^2]$ and $\hat{\sigma}_t^2 = \sum_{r=k+1}^p d_r/(p-k)$.

Low-rank compound Gaussian CD (1/2) [Mian et al., 2020]

Model and parameters

$$\mathbf{x}_i^t \sim \mathcal{CN}\left(\mathbf{0}, \tau_i^t \left(\mathbf{\Sigma}_k^t + \sigma_t^2 \mathbf{I}\right)\right)$$

where

$$\mathcal{L}_{\mathcal{CG}}\left(\{\mathbf{x}_i\}_{i=1}^n | \mathbf{\Sigma}, \{\tau_i\}_{i=1}^n\right) \propto \prod_{i=1}^n |\tau_i \mathbf{\Sigma}|^{-1} \exp\left(-\mathbf{x}_i^H (\tau_i \mathbf{\Sigma})^{-1} \mathbf{x}_i\right)$$

Hypothesis test

$$\begin{aligned} \mathrm{H}_0: \quad & \boldsymbol{\theta}_0 && = \left\{\mathbf{\Sigma}_k^0, \sigma_0^2, \{\tau_i^0\}_{i=1}^n\right\} \\ \mathrm{H}_1: \quad & \{\boldsymbol{\theta}_t\}_{t=1}^T && = \left\{\mathbf{\Sigma}_k^t, \sigma_t^2, \{\tau_i^t\}_{i=1}^n\right\}_{t=1}^T \end{aligned}$$

Low-rank compound Gaussian CD (2/2) [Mian et al., 2020]

Evaluating Λ_{LRCG} requires to compute:

$$\hat{\Lambda}_{\text{LRCG}} = \frac{\mathcal{L}_{LRCG}^{\text{H}_1}\left(\{\{\mathbf{x}_i^t\}_{i=1}^n\}_{t=1}^T \,|\hat{\boldsymbol{\theta}}_{\text{LRCG}}^{\text{H}_1}\right)}{\mathcal{L}_{LRCG}^{\text{H}_0}\left(\{\{\mathbf{x}_i^t\}_{i=1}^n\}_{t=1}^T \,|\hat{\boldsymbol{\theta}}_{\text{LRCG}}^{\text{H}_0}\right)},$$

where $\mathcal{L}_{LRCG}^{\text{H}_0}$ and $\mathcal{L}_{LRCG}^{\text{H}_1}$ are the likelihood under H_0 and H_1, and where

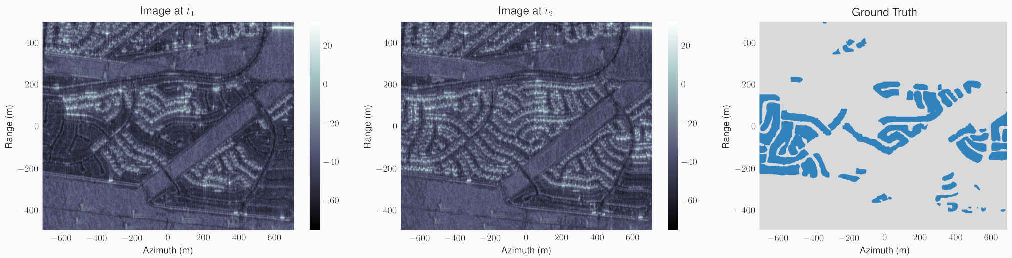
$$\hat{\boldsymbol{\theta}}_{\text{LRCG}}^{\text{H}_0} = \left\{\hat{\boldsymbol{\Sigma}}_k^0, \hat{\sigma}_0^2, \{\hat{\tau}_i^0\}_{i=1}^n\right\},$$
$$\hat{\boldsymbol{\theta}}_{\text{LRCG}}^{\text{H}_1} = \left\{\hat{\boldsymbol{\Sigma}}_k^t, \hat{\sigma}_t^2, \{\hat{\tau}_i^t\}_{i=1}^n\right\}_{t=1}^T$$

are the MLE under H_0 and H_1, respectively $\rightarrow$ evaluated with MM algorithm!

UAVSAR scene 1

Description

- Polarimetric data $\longrightarrow$ wavelet decomp. [Mian et al., 2017] $\longrightarrow$ $p = 12$ dim. pixels
- CD ground truth from [Nascimento et al., 2019]



Recall of the considered CD methods

Gaussian

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

Low-rank Gaussian

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

Compound-Gaussian

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

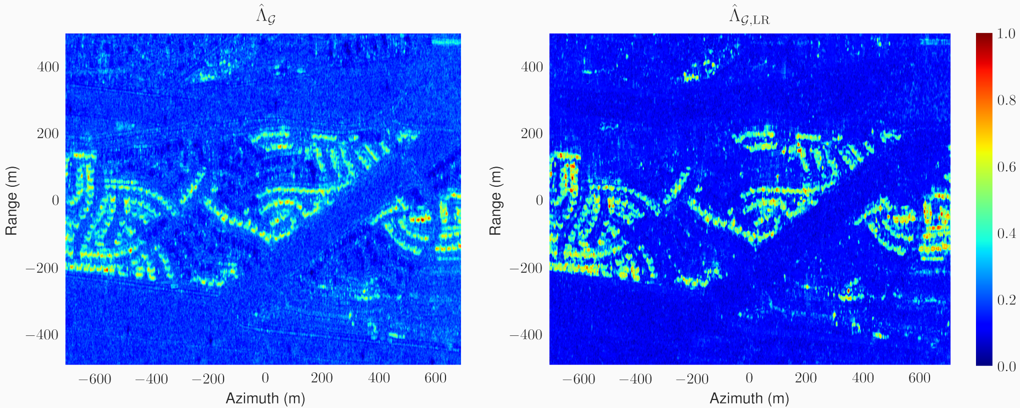
Low-rank Compound-Gaussian

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

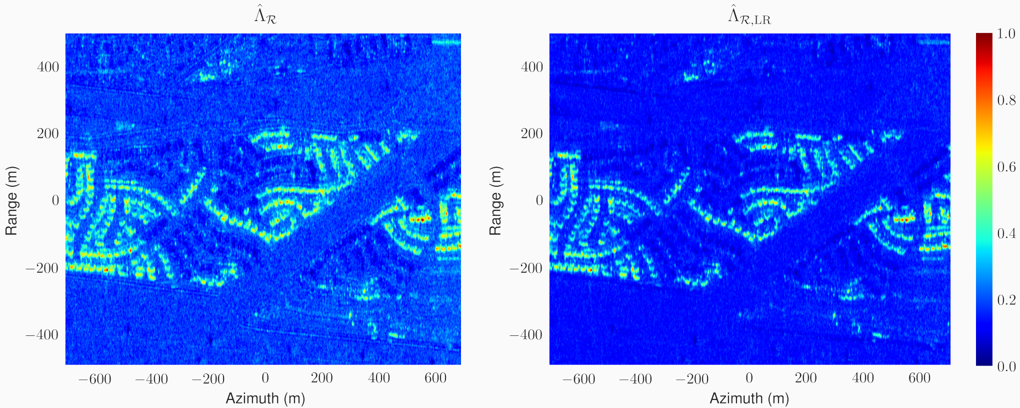
Variations on side parameters

Rank k and pre-estimated noise floor σ^2 detailed in [Mian et al., 2020]

Detectors output with a 5×5 sliding windows



Detectors output with a 5×5 sliding windows



Performance curves

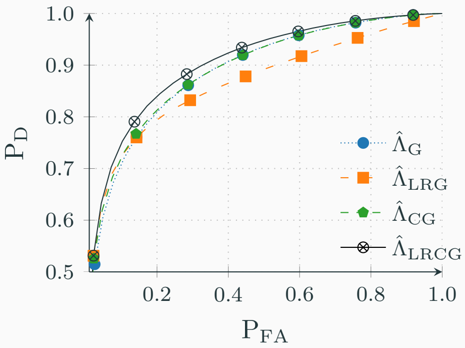


Figure 9: ROC with $(p = 12, N = 25, R = 3)$

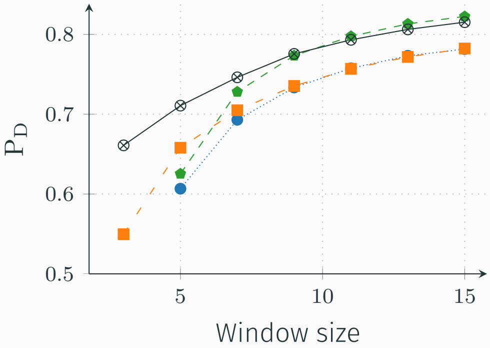


Figure 10: P_D versus window size $(P_{FA} = 5\%)$

Conclusions on CD for SAR-ITS

Conclusions

- Structures can be handled in robust models
- Improved performance and reduced window size

Perspectives

- Rank estimation strategies [Stoica and Selen, 2004, Terreaux et al., 2018]
- CFAR test statistic in Low-rank ?
  $\rightarrow$ Random Matrix theory correction [Vallet et al., 2019].
- Testing specific variations [Ben Abdallah et al., 2019]
- Sequential testing [Bouchard et al., 2020b]
- Clustering for time-series [Petitjean et al., 2012]

Final big picture

Part 1

- Detection through covariance
- 1/2-step procedures
- Gaussian framework

Part 2

- Robust framework
- Structured parameters
- Applications to SAR-ITS

Generic tools from this presentation

- Statistical (change) detection framework: “features and distances”
- CES models and robust covariance matrix estimation
- CD: Optimization methods (MM, Riemannian, (g-)convexity)

References i

- Andrews, D. F. and Mallows, C. L. (1974).
  Scale mixtures of normal distributions.
  Journal of the Royal Statistical Society. Series B (Methodological), 36(1):99–102.
- Ben Abdallah, R., Mian, A., Breloy, A., Taylor, A., El Korso, M. N., and Lautru, D. (2019).
  Detection methods based on structured covariance matrices for multivariate sar images processing.
  IEEE Geoscience and Remote Sensing Letters, 16(7):1160–1164.
- Besson, O. and Abramovich, Y. I. (2013).
  On the fisher information matrix for multivariate elliptically contoured distributions.
  IEEE Signal Processing Letters, 20(11):1130–1133.

Bouchard, F., Breloy, A., Ginolhac, G., Renaux, A., and Pascal, F. (2020a).
A riemannian framework for low-rank structured elliptical models.
arXiv preprint arXiv:2001.01141.

Bouchard, F., Mian, A., Zhou, J., Said, S., Ginolhac, G., and Berthoumieu, Y. (2020b).
Riemannian geometry for compound gaussian distributions: Application to recursive change detection.
Signal Processing, 176:107716.

Breloy, A., Ginolhac, G., Renaux, A., and Bouchard, F. (2019).
Intrinsic cramér–rao bounds for scatter and shape matrices estimation in ces distributions.
IEEE Signal Processing Letters, 26(2):262–266.

- Breloy, A., Ollila, E., and Pascal, F. (2019).
  Spectral shrinkage of tyler's m-estimator of covariance matrix.
  In *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 535–538. IEEE.
- Breloy, A., Sun, Y., Babu, P., Ginolhac, G., and Palomar, D. P. (2016).
  Robust rank constrained kronecker covariance matrix estimation.
  In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 810–814. IEEE.
- Couillet, R., Pascal, F., and Silverstein, J. W. (2015).
  The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples.
  Journal of Multivariate Analysis, 139:56–78.

Drašković, G. and Pascal, F. (2018).
New insights into the statistical properties of m-estimators.
IEEE Transactions on Signal Processing, 66(16):4253–4263.

Duembgen, L. and Tyler, D. E. (2016).
Geodesic convexity and regularized scatter estimators.
arXiv preprint arXiv:1607.05455.

Greco, M. and Gini, F. (2013).
Cramér-rao lower bounds on covariance matrix estimation for complex elliptically symmetric distributions.
IEEE Transactions on Signal Processing, 61(24):6401–6409.

References v

Hunter, D. R. and Lange, K. (2004).
A tutorial on mm algorithms.
The American Statistician, 58(1):30–37.

Maronna, R. A. (1976).
Robust m-estimators of multivariate location and scatter.
Ann. Statist., 4(1):51–67.

Meriaux, B., Ren, C., El Korso, M. N., Breloy, A., and Forster, P. (2019).
Robust estimation of structured scatter matrices in (mis) matched models.
Signal Processing, 165:163–174.

Mian, A., Collas, A., Breloy, A., Ginolhac, G., and Ovarlez, J. (2020).
Robust low-rank change detection for multivariate sar image time series.
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13:3545–3556.

Mian, A., Ginolhac, G., Ovarlez, J., and Atto, A. M. (2019a).
New robust statistics for change detection in time series of multivariate sar images.
IEEE Transactions on Signal Processing, 67(2):520–534.

Mian, A., Ovarlez, J., Atto, A. M., and Ginolhac, G. (2019b).
Design of new wavelet packets adapted to high-resolution sar images with an application to target detection.
IEEE Transactions on Geoscience and Remote Sensing, 57(6):3919–3932.

Mian, A., Ovarlez, J. P., Ginolhac, G., and Atto, A. (2017).
Multivariate change detection on high resolution monovariate sar image using linear time-frequency analysis.
In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 1942–1946.

Mian, A., Ovarlez, J.-P., Ginolhac, G., and Atto, A. M. (2017).
Multivariate change detection on high resolution monovariate SAR image using linear Time-Frequency analysis.
In *25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece.

Nascimento, A. D. C., Frery, A. C., and Cintra, R. J. (2019).
Detecting changes in fully polarimetric SAR imagery with statistical information theory.
IEEE Transactions on Geoscience and Remote Sensing, 57(3):1380–1392.

References viii

Ollila, E. and Tyler, D. E. (2014).
Regularized m-estimators of scatter matrix.
IEEE Transactions on Signal Processing, 62(22):6059–6070.

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

Pascal, F., Chitour, Y., Ovarlez, J.-P., Forster, P., and Larzabal, P. (2008).
Covariance structure maximum-likelihood estimates in compound Gaussian noise: existence and algorithm analysis.
Signal Processing, IEEE Transactions on, 56(1):34–48.

Petitjean, F., Inglada, J., and Gançarski, P. (2012).
Satellite image time series analysis under time warping.
IEEE transactions on geoscience and remote sensing, 50(8):3081–3095.

Ratha, D., De, S., Celik, T., and Bhattacharya, A. (2017).
Change Detection in Polarimetric SAR Images Using a Geodesic Distance Between Scattering Mechanisms.
IEEE Geoscience and Remote Sensing Letters, 14(7):1066–1070.

Soloveychik, I., Trushin, D., and Wiesel, A. (2015).
Group symmetric robust covariance estimation.
IEEE Transactions on Signal Processing, 64(1):244–257.

Soloveychik, I. and Wiesel, A. (2014).
Tyler's covariance matrix estimator in elliptical models with convex structure.
IEEE Transactions on Signal Processing, 62(20):5251–5259.

Soloveychik, I. and Wiesel, A. (2015).
Performance analysis of Tyler's covariance estimator.
Signal Processing, IEEE Transactions on, 63(2):418–426.

Stoica, P. and Selen, Y. (2004).
Model-order selection: a review of information criterion rules.
IEEE Signal Processing Magazine, 21(4):36–47.

Sun, Y., Babu, P., and Palomar, D. P. (2016a).
Majorization-minimization algorithms in signal processing, communications, and machine learning.
IEEE Transactions on Signal Processing, 65(3):794–816.

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

Sun, Y., Breloy, A., Babu, P., Palomar, D. P., Pascal, F., and Ginolhac, G. (2015).
Low-complexity algorithms for low rank clutter parameters estimation in radar systems.
IEEE Transactions on Signal Processing, 64(8):1986–1998.

Terreaux, E., Ovarlez, J., and Pascal, F. (2018).
A toeplitz-tyler estimation of the model order in large dimensional regime.
In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4489–4493.

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

Tyler, D. (1987).
A distribution-free m-estimator of multivariate scatter.
The Annals of Statistics, 15(1):234–251.

Vallet, P., Ginolhac, G., Pascal, F., and Forster, P. (2019).
An improved low rank detector in the high dimensional regime.
In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5336–5340.

Wiesel, A. (2011).
Unified framework to regularized covariance estimation in scaled gaussian models.
IEEE Transactions on Signal Processing, 60(1):29–38.

Wiesel, A. (2012).
Geodesic convexity and covariance estimation.
IEEE transactions on signal processing, 60(12):6182–6189.

Wiesel, A. and Zhang, T. (2015).
Structured robust covariance estimation.
Now Foundations and Trends.

Yao, K. (1973).
A Representation Theorem and its Applications to Spherically Invariant Random Processes.
Information Theory, IEEE Transactions on, 19:600–608.

Zhang, T., Cheng, X., and Singer, A. (2014).
Marchenko-Pastur Law for Tyler's and Maronna's M-estimators.
arXiv preprint.

Zhang, T., Wiesel, A., and Greco, M. S. (2013).
Multivariate generalized gaussian distribution: Convexity and graphical models.
IEEE Transactions on Signal Processing, 61(16):4141–4148.