Riemannian geometry for Compound Gaussian distributions: application to recursive change detection

Florent Bouchard[a], Ammar Mian[b], Jialun Zhou[c], Salem Said[c], Guillaume Ginolhac[a], Yannick Berthoumieu[c]

[a] *LISTIC, University Savoie Mont-Blanc, France*
[b] *SONDRA, CentraleSupelec, France*
[c] *IMS, University of Bordeaux, CNRS, France*

Abstract

A new Riemannian geometry for the zero-mean Compound Gaussian distribution with deterministic textures is proposed. In particular, the Fisher information metric (up to a factor) is obtained, along with corresponding geodesics and distance function. This new geometry is applied on a change detection problem on Multivariate Image Times Series: a recursive approach based on Riemannian optimization is developed. As shown on simulated data, it allows to reach optimal performance while being computationally more efficient.

Keywords: Riemaniann geometry and optimization, covariance matrix estimation, compound Gaussian distribution, change detection.

1. Introduction

Covariance matrix is an important topic in signal and image processing. When data are Gaussian distributed, the Maximum Likelihood Estimator (MLE) is the well known Sample Covariance Matrix (SCM). However, this

*Corresponding author
Email address: florent.bouchard@univ-smb.fr (Florent Bouchard)

estimator features poor performance when data follow a more heavy-tailed distribution. In such a case, it is interesting to model the data with a Complex Elliptically Symmetric (CES) distribution [1] and to employ M-estimators [2, 3, 4] for covariance estimation. In this paper, we limit ourselves to the Compound Gaussian (CG) distribution [5, 6], which is a CES sub-family. Its stochastic representation consists in a Gaussian vector multiplied by a positive scalar, called texture. For instance, this family fits well RADAR empirical data [7, 8].

It is possible to develop change detection algorithms for SAR Multivariate Image Times Series (MITS). Several approaches exist and those based on a test of equality of covariance matrices generally perform well. Moreover, they may have the interesting Constant False Alarm Rate (CFAR) property, in particular for Gaussian data [9, 10, 11]. For the CG distribution, the Generalized Likelihood Ratio Test (GLRT) is derived in [12]. This detector exhibits very good performance when data are not drawn from a Gaussian distribution. However, when the number of images T of the MITS is large, the computational time becomes prohibitive for practical implementation. In this paper, a recursive implementation of this detector is proposed. Because of the form of the change detector, this implementation cannot be derived easily, for example by employing an arithmetic mean.

To solve the problem, a framework based on a recursive approach as proposed in [13] is developed and adapted to the CG distribution. This approach is based on Riemannian geometry which is more and more used in the covariance estimation domain [14, 15, 16, 17]. However to derive such as recursive algorithm, the Riemannian geometry of the CG distribution has to be considered, which, to the best of our knowledge, has not been done previously. Hence, the main contribution of this paper consists in

deriving a well-suited Riemannian geometry for the distribution of interest, *i.e.* metric, geodesics, distance. It relies on the Fisher information metric (up to a factor) of the CG distribution; see *e.g.* [18, 19] for Gaussian and CES cases. In addition, the Riemannian gradient to recursively estimate the CG parameters of a MITS and the corresponding Intrinsic Cramér Rao Bound (ICRB) [18] are provided. Finally, the proposed method is validated on simulated data.

2. Data Model

Let us consider n i.i.d. samples $\{\mathbf{x}_i^{(t)}\}_{i\in[\![1,n]\!]}$ where each data $\mathbf{x}_i^{(t)} \in \mathbb{C}^p$. For example, these data may come from a spatial neighborhood of polarimetric channels (in this case $p = 3$) and/or spectro-angular diversity (see [20] for more details) in a SAR image. The MITS is composed of T sets of $\{\mathbf{x}_i^{(t)}\}_{i\in[\![1,n]\!]}$. In this MITS, even though these data follow the same statistical distribution, their parameters might change with t. From this MITS, we want to detect these changes by comparing the parameters of the distribution, denoted $\theta^{(t)}$. The change detection problem can be written as:

$$\begin{cases} \mathrm{H}_0 : \theta^{(1)} = \theta^{(2)} = ... = \theta^{(T)} = \theta^{(0)} \\ \mathrm{H}_1 : \exists (t, t') \in [\![1, T]\!]^2, \theta^{(t)} \neq \theta^{(t')} \end{cases} \tag{1}$$

As shown in [12], to reach good performance, it is important that the parameters capture both the power and the correlations of the data. To ensure this, we propose to use the CG distribution [5, 6] (also referred to as a mixture of scaled Gaussian). This model corresponds to a Gaussian one, where each realization $\mathbf{x}_i^{(t)} \in \mathbb{C}^p$ is scaled by a local power factor $\tau_i^{(t)}$ referred to as

texture sample (assumed unknown deterministic in this work):

$$\mathbf{x}_i^{(t)} \sim \mathcal{CN}(\mathbf{0}, \tau_i^{(t)} \mathbf{\Sigma}^{(t)}) \tag{2}$$

Therefore for this statistical modelling, we have $\theta^{(t)} = (\boldsymbol{\tau}^{(t)}, \mathbf{\Sigma}^{(t)})$ where $\boldsymbol{\tau}^{(t)} = (\tau_1^{(t)}, \ldots, \tau_n^{(t)})^T$. For the parameters to be identifiable, a constraint on the covariance $\mathbf{\Sigma}^{(t)}$ is needed. Most often, a trace constraint $\mathrm{tr}(\mathbf{\Sigma}^{(t)}) = p$ is applied. However, from a geometrical point of view, it is not the best choice. In the following, we choose the unitary determinant normalization, advocated in [21] because it allows to decorrelate the estimation of textures and covariance matrix. In this paper, we further show that it yields tremendous simplifications in the scaled Fisher information metric. Thus, $\mathbf{\Sigma}^{(t)}$ belongs to

$$\mathcal{SH}_p^{++} = \left\{ \mathbf{\Sigma} \in \mathcal{H}_p^{++} : \ |\mathbf{\Sigma}| = 1 \right\}, \tag{3}$$

where $\mathcal{H}_p^{++}$ is the manifold of $p \times p$ positive definite matrices.

In [12], the GLRT for the CG model is derived and the following detector is obtained:

$$\hat{\Lambda}_{CG}^{(T)} = \frac{\left|\hat{\mathbf{\Sigma}}_0^{(T)}\right|^{Tn}}{\prod_{t=1}^{T} \left|\hat{\mathbf{\Sigma}}_{Tyl}^{(t)}\right|^{n}} \prod_{i=1}^{i=n} \frac{\left(\sum_{t=1}^{T} \hat{\tau}_{i,0}^{(t)}\right)^{Tp}}{\prod_{t=1}^{T} \left(\hat{\tau}_i^{(t)}\right)^{p}} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda, \tag{4}$$

where $\hat{\mathbf{\Sigma}}_{Tyl}^{(t)}$ and $\hat{\tau}_i^{(t)}$ are provided by the classical Tyler's estimator [22, 23]:

$$\hat{\mathbf{\Sigma}}_{Tyl}^{(t)} = \frac{p}{n} \sum_{i=1}^{n} \frac{\mathbf{x}_i^{(t)} \mathbf{x}_i^{(t)H}}{\mathbf{x}_i^{(t)H} (\hat{\mathbf{\Sigma}}_{Tyl}^{(t)})^{-1} \mathbf{x}_i^{(t)}} \quad \text{and} \quad \hat{\tau}_i^{(t)} = \frac{\mathbf{x}_i^{(t)H} (\hat{\mathbf{\Sigma}}_{Tyl}^{(t)})^{-1} \mathbf{x}_i^{(t)}}{p}; \tag{5}$$

$\hat{\mathbf{\Sigma}}_0^{(T)}$ and $\hat{\tau}_{i,0}^{(T)}$ are the MLE of the covariance matrix and the textures under

the null hypothesis H_0:

$$\hat{\mathbf{\Sigma}}_0^{(T)} = \frac{p}{n}\sum_{i=1}^{n}\frac{\sum_{t=1}^{T}\mathbf{x}_i^{(t)}\mathbf{x}_i^{(t)H}}{\sum_{t=1}^{T}\mathbf{x}_i^{(t)H}(\hat{\mathbf{\Sigma}}_0^{(T)})^{-1}\mathbf{x}_i^{(t)}} \quad \text{and} \quad \hat{\tau}_{i,0}^{(T)} = \sum_{t=1}^{T}\frac{\mathbf{x}_i^{(t)H}(\hat{\mathbf{\Sigma}}_0^{(T)})^{-1}\mathbf{x}_i^{(t)}}{Tp}. \tag{6}$$

This detector features interesting CFAR properties and exhibits better performances when data follow a CG distribution. Unfortunately, it suffers a large complexity, in particular as T grows. Moreover, when a new dataset $\{\mathbf{x}_i^{(T+1)}\}$ occurs, it is impossible to compute the new detector $\hat{\Lambda}_{CG}^{(T+1)}$ directly from $\hat{\Lambda}_{CG}^{(T)}$ because:

$$\hat{\mathbf{\Sigma}}_0^{(T+1)} \neq \frac{T\hat{\mathbf{\Sigma}}_0^{(T)} + \hat{\mathbf{\Sigma}}_{Tyl}^{(T+1)}}{T+1} \tag{7}$$

To avoid the computation of $\hat{\mathbf{\Sigma}}_0^{(T+1)}$ with all previous data, an original recursive approach based on Riemannian optimization is proposed.

3. Riemannian geometry of the compound Gaussian distribution

To simplify notations, the superscript $^{(t)}$ is omitted in this section. In the following, $\boldsymbol{\tau} = [\tau_1 \dots \tau_n]^T$, $\theta = (\mathbf{\Sigma}, \boldsymbol{\tau})$, $\xi = (\xi_{\mathbf{\Sigma}}, \xi_{\boldsymbol{\tau}})$ and $\eta = (\eta_{\mathbf{\Sigma}}, \eta_{\boldsymbol{\tau}})$. The parameter θ of the CG distribution lies in the manifold $\mathcal{M}_{p,n} = \mathcal{SH}_p^{++} \times \mathbb{R}_{++}^n$. Since this is the product of two manifolds, $\mathcal{M}_{p,n}$ is also a manifold (see *e.g.* [24] for details). Its tangent space $T_\theta\mathcal{M}_{p,n}$ at θ is $T_{\mathbf{\Sigma}}\mathcal{SH}_p^{++} \times T_{\boldsymbol{\tau}}\mathbb{R}_{++}^n$, where $T_{\mathbf{\Sigma}}\mathcal{SH}_p^{++}$:

$$T_{\mathbf{\Sigma}}\mathcal{SH}_p^{++} = \{\xi_{\mathbf{\Sigma}} \in \mathcal{H}_p : \ \mathrm{tr}(\mathbf{\Sigma}^{-1}\xi_{\mathbf{\Sigma}}) = 0\} \tag{8}$$

($\mathcal{H}_p$ denotes the space of $p \times p$ Hermitian matrices); and $T_{\boldsymbol{\tau}}\mathbb{R}^n_{++}$ is identified to $\mathbb{R}^n$.

To turn $\mathcal{M}_{p,n}$ into a Riemannian manifold, it must be equiped with a Riemannian metric. The most natural choice in our case is to consider the Fisher information metric (up to a factor) on $\mathcal{M}_{p,n}$ associated with the CG distribution. It is given in the following proposition.

Proposition 3.1 (Scaled Fisher information metric)**.** *The scaled Fisher metric of the CG distribution on $\mathcal{M}_{p,n}$ is defined, for $\theta \in \mathcal{M}_{p,n}$ and $\xi, \eta \in T_\theta \mathcal{M}_{p,n}$, by*

$$\langle \xi, \eta \rangle_\theta^{\mathcal{M}_{p,n}} = \frac{1}{p} \langle \xi_{\boldsymbol{\Sigma}}, \eta_{\boldsymbol{\Sigma}} \rangle_{\boldsymbol{\Sigma}}^{\mathcal{H}_p^{++}} + \frac{1}{n} \langle \xi_{\boldsymbol{\tau}}, \eta_{\boldsymbol{\tau}} \rangle_{\boldsymbol{\tau}}^{\mathbb{R}^n_{++}},$$

with $\langle \xi_{\boldsymbol{\Sigma}}, \eta_{\boldsymbol{\Sigma}} \rangle_{\boldsymbol{\Sigma}}^{\mathcal{H}_p^{++}} = tr(\boldsymbol{\Sigma}^{-1} \xi_{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} \eta_{\boldsymbol{\Sigma}})$ and $\langle \xi_{\boldsymbol{\tau}}, \eta_{\boldsymbol{\tau}} \rangle_{\boldsymbol{\tau}}^{\mathbb{R}^n_{++}} = (\xi_{\boldsymbol{\tau}} \odot \boldsymbol{\tau}^{\odot -1})^T \eta_{\boldsymbol{\tau}} \odot \boldsymbol{\tau}^{\odot -1}$, where $\odot$ and $\cdot^{\odot -1}$ denote elementwise product and inversion, respectively.

Proof. The log-likelihood L_{CG} on $\mathcal{M}_{p,n}$ for θ is

$$L_{\mathrm{CG}}(\theta) = \sum_i L_{\mathrm{G}}(\tau_i \boldsymbol{\Sigma}) = \sum_i L_{\mathrm{G}} \circ \varphi_i(\theta), \tag{9}$$

where L_{G} is the log-likelihood for the Gaussian distribution, see *e.g.* [18]; and $\varphi_i(\theta) = \tau_i \boldsymbol{\Sigma}$. By definition and [18, Theorem 1],

$$\begin{aligned}
\langle \xi, \eta \rangle_\theta^{\mathcal{M}_{p,n}} &= \mathbb{E}\left[\mathrm{D}\, L_{\mathrm{CG}}(\theta)[\xi]\, \mathrm{D}\, L_{\mathrm{CG}}(\theta)[\eta]\right] = -\mathbb{E}\left[\mathrm{D}^2\, L_{\mathrm{CG}}(\theta)[\xi, \eta]\right] \\
&= -\textstyle\sum_i \mathbb{E}\left[\mathrm{D}^2\, L_{\mathrm{G}} \circ \varphi_i(\theta)[\xi, \eta]\right] \\
&= \textstyle\sum_i \mathbb{E}\left[\mathrm{D}\, L_{\mathrm{G}} \circ \varphi_i(\theta)[\xi]\, \mathrm{D}\, L_{\mathrm{G}} \circ \varphi_i(\theta)[\eta]\right] \\
&= \textstyle\sum_i \langle \mathrm{D}\, \varphi_i(\theta)[\xi], \mathrm{D}\, \varphi_i(\theta)[\eta] \rangle_{\varphi_i(\theta)}^{\mathcal{H}_p^{++}},
\end{aligned}$$

where $\mathrm{D}\, \varphi_i(\theta)[\xi] = \xi_{\boldsymbol{\tau} i} \boldsymbol{\Sigma} + \tau_i \xi_{\boldsymbol{\Sigma}}$ is the directional derivative of φ_i, with $\xi_{\boldsymbol{\tau} i}$

the i^{th} element of $\boldsymbol{\xi}_{\boldsymbol{\tau}}$. Basic manipulations yield, up to a factor,

$$\langle \xi, \eta \rangle_{\theta}^{\mathcal{M}_{p,n}} = \frac{1}{p} \langle \xi_{\boldsymbol{\Sigma}}, \eta_{\boldsymbol{\Sigma}} \rangle_{\boldsymbol{\Sigma}}^{\mathcal{H}_p^{++}} + \frac{1}{n} \langle \xi_{\boldsymbol{\tau}}, \eta_{\boldsymbol{\tau}} \rangle_{\boldsymbol{\tau}}^{\mathbb{R}^n_{++}} + \frac{1}{np} \operatorname{tr}(\boldsymbol{\Sigma}^{-1} \xi_{\boldsymbol{\Sigma}}) (\eta_{\boldsymbol{\tau}} \odot \boldsymbol{\tau}^{-1})^T \mathbf{1}_n + \frac{1}{np} \operatorname{tr}(\boldsymbol{\Sigma}^{-1} \eta_{\boldsymbol{\Sigma}}) (\xi_{\boldsymbol{\tau}} \odot \boldsymbol{\tau}^{-1})^T \mathbf{1}_n.$$

Since $\xi_{\boldsymbol{\Sigma}}, \eta_{\boldsymbol{\Sigma}} \in T_{\boldsymbol{\Sigma}} \mathcal{SH}_p^{++}$, we have $\operatorname{tr}(\boldsymbol{\Sigma}^{-1} \xi_{\boldsymbol{\Sigma}}) = \operatorname{tr}(\boldsymbol{\Sigma}^{-1} \eta_{\boldsymbol{\Sigma}}) = 0$, which concludes the proof. □

In the following proposition, the geodesics and Riemannian distance on $\mathcal{M}_{p,n}$ associated with the scaled Fisher information metric $\langle \cdot, \cdot \rangle^{\mathcal{M}_{p,n}}$ of the CG distribution are provided. These geometrical objects are sufficient to perform Riemannian optimization and to measure and bound estimation errors.

Proposition 3.2 (Geodesics and Riemannian distance)**.** *The geodesic on $\mathcal{M}_{p,n}$ is $\gamma^{\mathcal{M}_{p,n}}(t) = (\gamma^{\mathcal{SH}_p^{++}}(t), \gamma^{\mathbb{R}^n_{++}}(t))$. If $\gamma^{\mathcal{M}_{p,n}}(0) = \theta$ and $\dot{\gamma}^{\mathcal{M}_{p,n}}(0) = \xi$,*

$$\gamma^{\mathcal{SH}_p^{++}}(t) = \boldsymbol{\Sigma} \operatorname{expm}(t \boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}_{\boldsymbol{\Sigma}}) \quad \text{and} \quad \gamma^{\mathbb{R}^n_{++}}(t) = \boldsymbol{\tau} \odot \exp(t \boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\xi}_{\boldsymbol{\tau}}).$$

If $\gamma^{\mathcal{M}_{p,n}}(0) = \theta_0$ and $\gamma^{\mathcal{M}_{p,n}}(1) = \theta_1$,

$$\gamma^{\mathcal{SH}_p^{++}}(t) = \boldsymbol{\Sigma}_0^{1/2} (\boldsymbol{\Sigma}_0^{-1/2} \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_0^{-1/2})^t \boldsymbol{\Sigma}_0^{1/2} \quad \text{and} \quad \gamma^{\mathbb{R}^n_{++}}(t) = \boldsymbol{\tau}_0^{\odot 1-t} \odot \boldsymbol{\tau}_1^{\odot t}.$$

It follows that the Riemannian distance on $\mathcal{M}_{p,n}$ corresponding to the scaled Fisher metric of proposition 3.1 is

$$\delta^2_{\mathcal{M}_{p,n}}(\theta_0, \theta_1) = \frac{1}{p} \delta^2_{\mathcal{H}_p^{++}}(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) + \frac{1}{n} \delta^2_{\mathbb{R}^n_{++}}(\boldsymbol{\tau}_0, \boldsymbol{\tau}_1),$$

where $\delta^2_{\mathcal{H}_p^{++}}(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) = \| \operatorname{logm}(\boldsymbol{\Sigma}_0^{-1/2} \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_0^{-1/2}) \|_2^2$ and $\delta^2_{\mathbb{R}^n_{++}}(\boldsymbol{\tau}_0, \boldsymbol{\tau}_1) = \| \log(\boldsymbol{\tau}_0^{-1} \odot$

$\boldsymbol{\tau}_1)\|_2^2$.

Proof. The geodesics $\gamma^{\mathcal{SH}_p^{++}}(t)$ and $\gamma^{\mathbb{R}_{++}^n}(t)$ are the geodesics on $\mathcal{SH}_p^{++}$ and $\mathbb{R}_{++}^n$ equiped with $\langle\cdot,\cdot\rangle_{\cdot}^{\mathcal{H}_p^{++}}$ and $\langle\cdot,\cdot\rangle_{\cdot}^{\mathbb{R}_{++}^n}$, respectively. Therefore, by definition of $\langle\cdot,\cdot\rangle_{\cdot}^{\mathcal{M}_{p,n}}$ and from the properties of product manifolds, $\gamma^{\mathcal{M}_{p,n}}$ is the geodesic on $\mathcal{M}_{p,n}$. Similarly, $\delta^2_{\mathcal{H}_p^{++}}$ and $\delta^2_{\mathbb{R}_{++}^n}$ are the Riemannian distances associated with $\langle\cdot,\cdot\rangle_{\cdot}^{\mathcal{H}_p^{++}}$ and $\langle\cdot,\cdot\rangle_{\cdot}^{\mathbb{R}_{++}^n}$. Thus, by definition of $\langle\cdot,\cdot\rangle_{\cdot}^{\mathcal{M}_{p,n}}$, $\delta^2_{\mathcal{M}_{p,n}}$ is the associated Riemannian distance on $\mathcal{M}_{p,n}$. □

4. Application to recursive change detection

Given a new data-set of n samples at $t+1$ $\{\mathbf{x}_i^{(t+1)}\}_{i\in[\![1,n]\!]}$, to obtain the CG change detector $\hat{\Lambda}_{CG}^{(t+1)}$ defined in (4), one needs to compute: $\hat{\theta}_{Tyl}^{(t+1)} = (\hat{\boldsymbol{\Sigma}}_{Tyl}^{(t+1)}, \hat{\boldsymbol{\tau}}_{Tyl}^{(t+1)})$ and $\hat{\theta}_0^{(t+1)} = (\hat{\boldsymbol{\Sigma}}_0^{(t+1)}, \hat{\boldsymbol{\tau}}_0^{(t+1)})$ defined in (5) and (6). The complexity of the computation of $\hat{\theta}_0^{(t+1)}$ with usual techniques is quite high. To solve this issue, a recursive implementation, obtained by exploiting the Riemannian derivation studied in [13], is proposed. To estimate $\hat{\theta}_0^{(t+1)}$, we only use the information provided by $\hat{\theta}_0^{(t)}$ and the log-likelihood of the new data $\{\mathbf{x}_i^{(t+1)}\}_i$, which is

$$L_{CG}^{(t+1)}(\theta) = \sum_i -p\log(\tau_i) - \frac{(\mathbf{x}_i^{(t+1)})^H \boldsymbol{\Sigma}^{-1}\mathbf{x}_i^{(t+1)}}{\tau_i}. \tag{10}$$

The recursive algorithm returning the sequence of estimates $\{\theta_0^{(t)}\}_t$ corresponding to the sequence of data $\{\mathbf{x}_i^{(t)}\}_{i,t}$ is given in Algorithm 1. This algorithm relies on: (i) the Riemannian exponential map $\exp_\theta^{\mathcal{M}_{p,n}} : T_\theta\mathcal{M}_{p,n} \to \mathcal{M}_{p,n}$, such that $\exp_\theta^{\mathcal{M}_{p,n}}(\xi) = \gamma^{\mathcal{M}_{p,n}}(1)$, where $\gamma^{\mathcal{M}_{p,n}}$ is defined in Proposition 3.2; (ii) the Riemannian gradient of $L_{CG}^{(t)}$, provided in Proposition 4.1.

Algorithm 1: Recursive estimation of CG parameters in $\mathcal{M}_{p,n}$

Input: $\{\boldsymbol{x}_i^{(t)}\}_{i,t}$, initialization $\theta^{(0)} \in \mathcal{M}_{p,n}$, initial stepsize $\alpha_0 > 0$
Output: $\{\theta^{(t)}\}_t$ in $\mathcal{M}_{p,n}$
for $t = 0$ ***to*** T **do**
  $\theta^{(t+1)} = \exp_{\theta^{(t)}}^{\mathcal{M}_{p,n}} \left(\frac{\alpha_0}{t+1} \operatorname{grad}_{\mathcal{M}_{p,n}} L_{CG}^{(t+1)}(\theta^{(t)}) \right)$

Proposition 4.1 (Gradient of the parameters of CG distribution)**.** *The Riemannian gradient* $\operatorname{grad}_{\mathcal{M}_{p,n}} L_{CG}^{(t)}(\theta)$ *at* $\theta \in \mathcal{M}_{p,n}$ *is*

$$\operatorname{grad}_{\mathcal{M}_{p,n}} L_{CG}^{(t)}(\theta) = \left(\sum_i \frac{p\, \mathbf{x}_i^{(t)} (\mathbf{x}_i^{(t)})^H - (\mathbf{x}_i^{(t)})^H \boldsymbol{\Sigma}^{-1} \mathbf{x}_i^{(t)}\, \boldsymbol{\Sigma}}{\tau_i}, n(\boldsymbol{a} - p\boldsymbol{\tau}) \right)$$

where $\boldsymbol{a} \in \mathbb{R}^n$*, whose* i^{th} *element is* $a_i = (\mathbf{x}_i^{(t)})^H \boldsymbol{\Sigma}^{-1} \mathbf{x}_i^{(t)}$.

Proof. By definition [24], for all $\xi \in T_\theta \mathcal{M}_{p,n}$, $\langle \operatorname{grad}_{\mathcal{M}_{p,n}} L_{CG}^{(t)}(\theta), \xi \rangle_\theta^{\mathcal{M}_{p,n}} = \mathrm{D}\, L_{CG}^{(t)}(\theta)[\xi]$. We have

$$\begin{aligned} \mathrm{D}\, L_{CG}^{(t)}(\theta)[\xi] &= \textstyle\sum_i \frac{(\mathbf{x}_i^{(t)})^H \boldsymbol{\Sigma}^{-1} \mathbf{x}_i^{(t)} - p\tau_i}{\tau_i^2} \xi_{\boldsymbol{\tau}\, i} + \frac{(\mathbf{x}_i^{(t)})^H \boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}_{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} \mathbf{x}_i^{(t)}}{\tau_i} \\ &= \tfrac{1}{n} \langle n(\boldsymbol{a} - p\boldsymbol{\tau}), \boldsymbol{\xi}_{\boldsymbol{\tau}} \rangle_{\boldsymbol{\tau}}^{\mathbb{R}^n_{++}} + \tfrac{1}{p} \langle p \textstyle\sum_i \frac{\mathbf{x}_i^{(t)} (\mathbf{x}_i^{(t)})^H}{\tau_i}, \boldsymbol{\xi}_{\boldsymbol{\Sigma}} \rangle_{\boldsymbol{\Sigma}}^{\mathcal{H}_p^{++}}. \end{aligned}$$

It remains to project $p \sum_i \frac{\mathbf{x}_i^{(t)} (\mathbf{x}_i^{(t)})^H}{\tau_i}$ on the tangent space $T_{\boldsymbol{\Sigma}} \mathcal{SH}_p^{++}$. This is achieved by using $P_{\boldsymbol{\Sigma}}^{\mathcal{SH}_p^{++}}(\boldsymbol{\xi}_{\boldsymbol{\Sigma}}) = \operatorname{herm}(\boldsymbol{\xi}_{\boldsymbol{\Sigma}}) - \frac{1}{p} \operatorname{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}_{\boldsymbol{\Sigma}}) \boldsymbol{\Sigma}$ (see *e.g.* [19]). One can check that it yields the proposed gradient. □

The Riemannian distance in Proposition 3.2 can be used to measure the error contained in an unbiased estimator $\hat{\theta}^{(T)}$ of the parameter $\theta^{(T)}$ corresponding to a MITS with T data. Exploiting the same framework as in [18, 19], the corresponding ICRB is provided in the following proposition.

Proposition 4.2 (ICRB)**.** *Given an unbiased estimator* $\hat{\theta}^{(T)}$ *of* $\theta^{(T)}$ *cor-*

responding to a MITS with T data, the ICRB corresponding to the error measured with the Riemannian distance in Proposition 3.2 is

$$\mathbb{E}[\delta^2_{\mathcal{M}_{p,n}}(\theta^{(T)}, \hat{\theta}^{(T)})] \leq \frac{p^2 - 1 + n}{Tpn}$$

Proof. By definition of $\langle \cdot, \cdot, \rangle^{\mathcal{M}_{p,n}}$ in Proposition 3.1, the Fisher information matrix is $\boldsymbol{F} = Tpn\boldsymbol{I}_{p^2-1+n}$. Thus, $\mathrm{tr}(\boldsymbol{F}^{-1}) = \frac{p^2-1+n}{Tpn}$, which is enough to conclude. □

5. Numerical simulations

Given T data, the performance of the CG change detector (4) under the null hypothesis greatly depends on the quality of the estimator $\theta_0^{(T)}$. In this numerical experiment, we compare the performance of the three following estimators:

- The MLE $\hat{\theta}_{mle}$, which features the best performance but is computationally expensive.
- The arithmetic mean $\hat{\theta}_{art}$, such that $\hat{\theta}_{art}^{(t+1)} = \frac{t\hat{\theta}_{art}^{(t)} + \hat{\theta}_{Tyl}^{(t+1)}}{t+1}$, where $\hat{\theta}_{Tyl}^{(t+1)}$ is Tyler's estimator (5) of $\{\mathbf{x}_i^{(t+1)}\}_i$.
- The recursive estimation $\hat{\theta}_{rec}$ proposed in Algorithm 1 with $\alpha_0 = 1/pn$.

Simulated data $\{\mathbf{x}_i^{(t)}\}_{i,t}$ of size $p = 10$, $n \in \{20, 50\}$, $T \in [\![1, 1000]\!]$ are drawn from a K-distribution. Textures $\boldsymbol{\tau}$ follow a Γ distribution with parameters $\alpha = \beta = 1$. The covariance matrix is generated as $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, where $\mathbf{U}$ is a random unitary matrix drawn from a normal distribution and $\boldsymbol{\Lambda}$ is a random diagonal positive definite matrix with unitary determinant drawn from a chi-squared distribution.

In Figure 1, we observe that, as expected, the MLE features the best performance and quickly gets close to the ICRB as T grows. The arithmetic mean has good performance for small values of T but reaches a minimal floor, thus displaying poor performance for large T. Finally, our proposed method works quite well: it reaches optimal performance as T grows. Further notice that a smaller amount of samples are needed to get close to the bound as n increases. Moreover, it has the smallest complexity as only one iteration is needed for each new incoming data. This property is shown in Figure 2 where the computation w.r.t. T for the three methods is plotted.

6. Conclusion

We have adapted a change detector derived for CG data in order to execute it recursively and greatly reduce the complexity of the calculation. This approach is based on Riemannian optimization which required the construction of geometry for CG distribution. Simulations have shown the interest of this new algorithm to reduce the complexity while maintaining good performance. In the future, this approach has to be tested on real SAR data.

Acknowledgment

This work was supported by ANR PHOENIX (ANR-15-CE23-0012) and ANR-ASTRID MARGARITA (ANR-17-ASTR-0015).

References

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

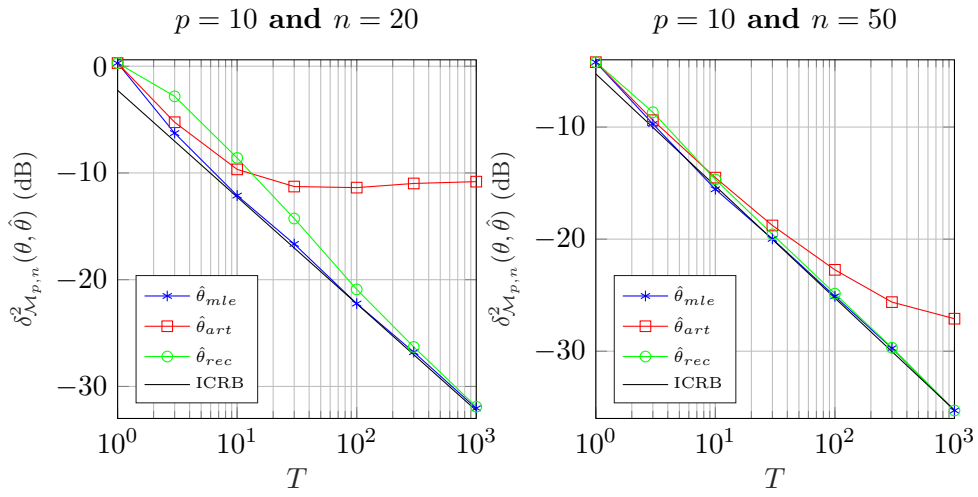

Figure 1: MSE $\delta^2_{\mathcal{M}_{p,n}}(\theta, \hat{\theta})$ as a function of T with $p = 10$, $n = 20$ (left) and $n = 50$ (right).

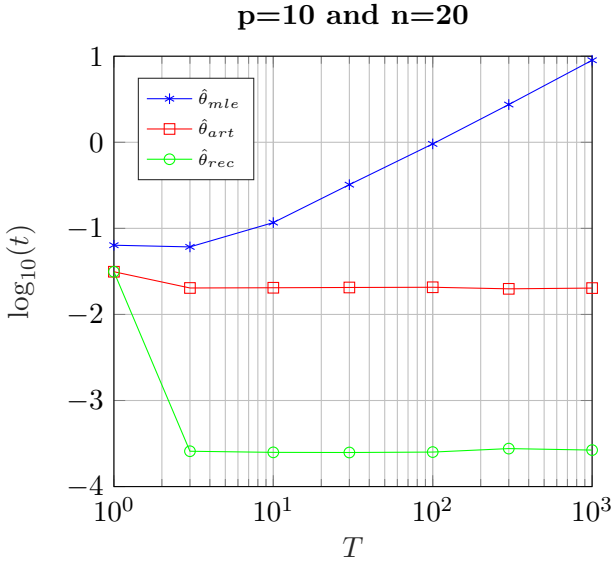

Figure 2: Time computation in log scale as a function of T with $p = 10$ and $n = 20$.

[2] E. Conte, A. D. Maio, G. Ricci, Recursive estimation of the covariance matrix of a compound-gaussian process and its application to adaptive cfar detection, IEEE Transactions on Signal Processing 50 (8) (2002)

1908–1915.

[3] Y. Abramovich, O. Besson, Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach-part 1: The over-sampled case, IEEE Transactions on Signal Processing 61 (23) (2013) 5807–5818.

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

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

[6] F. Gini, M. Greco, Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter, Signal Processing 82 (12) (2002) 1847–1859.

[7] E. Conte, M. Longo, Characterisation of radar clutter as a spherically invariant random process, IEE Proceedings Communications, Radar and Signal Processing 134 (2) (1987) 191–197.

[8] M. Greco, F. Gini, Statistical analysis of high-resolution SAR ground clutter data, IEEE Transactions on Geoscience and Remote Sensing 45 (3) (2007) 566–575.

[9] K. Conradsen, A. A. Nielsen, J. Schou, H. Skriver, A test statistic in the complex Wishart distribution and its application to change detection in

polarimetric SAR data, IEEE Transactions on Geoscience and Remote Sensing 41 (1) (2003) 4–19.

[10] L. M. Novak, Coherent change detection for multi-polarization SAR, in: Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers, 2005., 2005, pp. 568–573.

[11] D. Ciuonzo, V. Carotenuto, A. D. Maio, On multiple covariance equality testing with application to SAR change detection, IEEE Transactions on Signal Processing 65 (19) (2017) 5078–5091.

[12] A. Mian, G. Ginolhac, J.-P. Ovarlez, A. M. Atto, New robust statistics for change detection in time series of multivariate SAR images, IEEE Transactions on Signal Processing 67 (2) (2019) 520–534.

[13] J. Zhou, S. Said, Fast, asymptotically efficient, recursive estimation in a Riemannian manifold, Entropy 21 (10).

[14] F. Barbaresco, Information geometry of covariance matrix: Cartan-Siegel homogeneous bounded domains, Mostow/Berger fibration and frechet median, in: Matrix information geometry, Springer, Berlin, Heidelberg, Germany, 2013, pp. 199–255.

[15] M. Arnaudon, F. Barbaresco, L. Yang, Riemannian medians and means with applications to radar signal processing, IEEE Journal of Selected Topics in Signal Processing 7 (4) (2013) 595–604.

[16] G. Cui, N. Li, L. Pallotta, G. Foglia, L. Kong, Geometric barycenters for covariance estimation in compound-gaussian clutter, IET Radar, Sonar and Navigation 11 (3) (2017) 404–409.

[17] A. Aubry, A. D. Maio, L. Pallotta, A geometric approach to covariance matrix estimation and its applications to radar problems, IEEE Transactions on Signal Processing 66 (4) (2018) 907–922.

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

[19] A. Breloy, G. Ginolhac, A. Renaux, F. Bouchard, Intrinsic Cramér-Rao bounds for scatter and shape matrices estimation in CES distributions, IEEE Signal Processing Letters 26 (2) (2019) 262–266.

[20] A. Mian, J.-P. Ovarlez, A. M. Atto, G. Ginolhac, 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) (2019) 3919–3932.

[21] D. Paindaveine, A canonical definition of shape, Statistics & probability letters 78 (14) (2008) 2240–2247.

[22] D. E. Tyler, A distribution-free M-estimator of multivariate scatter, The Annals of Statistics 15 (1) (1987) 234–251.

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

[24] P.-A. Absil, R. Mahony, R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, Princeton, NJ, USA, 2008.