A TYLER-TYPE ESTIMATOR OF LOCATION AND SCATTER LEVERAGING RIEMANNIAN OPTIMIZATION

A. Collas[1], F. Bouchard[2], A. Breloy[3], C. Ren[1], G. Ginolhac[4], J.-P. Ovarlez[1,5]

[1]SONDRA, CentraleSupélec, Université Paris-Saclay
[2]CNRS, L2S, CentraleSupélec, Université Paris-Saclay
[3]LEME, Université Paris Nanterre
[4]LISTIC, Université Savoie Mont Blanc
[5]DEMR, ONERA, Université Paris-Saclay

ABSTRACT

We consider the problem of jointly estimating the location and scatter matrix of a Compound Gaussian distribution with unknown deterministic texture parameters. When the location is known, the Maximum Likelihood Estimator (MLE) of the scatter matrix corresponds to Tyler's M-estimator, which can be computed using fixed point iterations. However, when the location is unknown, the joint estimation problem remains challenging since the associated standard fixed-point procedure to evaluate the solution may often diverge. In this paper, we propose a stable algorithm based on Riemannian optimization for this problem. Finally, numerical simulations show the good performance and usefulness of the proposed algorithm.

Index Terms— Compound Gaussian, Robust location and scatter estimation, Riemannian optimization

1. INTRODUCTION

Many signal processing and machine learning tasks require estimates of the first and second order statistical moments of the sample set $\{x_i\}_{i=1}^n$ [1–4]. Classically, these parameters are estimated using the empirical mean and the sample covariance matrix (SCM). However, these estimates tend to perform poorly in the context of heavy-tailed distributions or when the set contains outliers, which motivates the use of robust estimation methods. Notably, M-estimators [5], raised increasing interest in the past decades (see e.g. [6]). These estimators are expressed through the two joint fixed point equations

$$\boldsymbol{\mu} = \Big(\sum_{i=1}^n u_1(t_i) \Big)^{-1} \sum_{i=1}^n u_1(t_i) \boldsymbol{x_i} \triangleq \mathcal{H}_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\Sigma} = \frac{1}{n} \sum_{i=1}^n u_2(t_i)(\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x_i} - \boldsymbol{\mu})^H \triangleq \mathcal{H}_{\boldsymbol{\Sigma}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

(1)

where $t_i \triangleq (\boldsymbol{x_i} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{x_i} - \boldsymbol{\mu})$, u_1 and u_2 are functions that respect Maronna's conditions[1] [5]. Under certain condi-

---

[1]Notice that [5] rather uses a formulation of (1) involving "$u_1(t_i)$" and "$u_2(t_i^2)$", with $t_i = \sqrt{(\boldsymbol{x_i} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{x_i} - \boldsymbol{\mu})}$. Without loss of general-

tions [5], these estimators can be computed with fixed-point iterations

$$\boldsymbol{\mu}_{k+1} = \mathcal{H}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$
$$\boldsymbol{\Sigma}_{k+1} = \mathcal{H}_{\boldsymbol{\Sigma}}(\boldsymbol{\mu}_{k+1}, \boldsymbol{\Sigma}_k)$$

(2)

that converge towards a unique solution satisfying (1). Interestingly, some M-estimators also appear as Maximum Likelihood Estimators when $u_1(t) = u_2(t)$ is linked to the p.d.f. of an elliptical distribution [6]. Expressing these estimators as the solution of an optimization problem drove a recent line of work leveraging optimization theory allowing, e.g., for generalizations to structured scatter matrix matrices [7–9] or regularized location and scatter matrix estimation [10].

In the context of scatter matrix estimation, Tyler's M-estimator [11] (obtained for $\boldsymbol{\mu} = \mathbf{0}$ and $u_2(t) = p/t$) is especially interesting thanks to its "distribution-free" properties over the elliptical distributions family. However this estimator cannot trivially be transposed to the case of joint mean-scatter matrix estimation by using $u_1(t) = u_2(t) = p/t$, that does not satisfy Maronna's conditions [5], and for which the fixed-point iterations (2) generally diverge. Thus, Tyler's M-estimator of the scatter matrix is usually applied on de-meaned data (where the mean is estimated in a prior step). A closely related estimator proposed by Tyler in [11] uses $u_1(t) = \sqrt{p/t}$ and $u_2(t) = p/t$, which yields converging fixed-point iterations in practice despite being a limit case of Maronna's conditions.

In this paper, we consider a statistical model for which a counterpart of Tyler's M-estimator (i.e., using $u_1(t) = u_2(t) = p/t$) for joint location-scatter matrix appears as MLE (specifically, non-centered compound Gaussian with unknown deterministic textures distribution). We then leverage the framework of Riemannian optimization in order to propose a gradient-type algorithm that converges to this solution, even when the fixed-point iterations (2) diverge. The performance of this approach is then illustrated in simulations.

---

ity, this paper uses the present notation in order to simplify some discussions.

2. MODEL

Let a set of n data points $\boldsymbol{x_i} \in \mathbb{C}^p$ distributed according to the following model

$$\boldsymbol{x_i} \underset{d}{=} \boldsymbol{\mu} + \sqrt{\tau_i}\,\boldsymbol{\Sigma}^{\frac{1}{2}}\,\boldsymbol{u_i}\,, \qquad (3)$$

where $\boldsymbol{u_i} \sim \mathbb{CN}(\boldsymbol{0}, \boldsymbol{I}_p)$ circular. The variables $\boldsymbol{\mu} \in \mathbb{C}^p$ and $\boldsymbol{\Sigma} \in \mathcal{H}_p^{++}$ (set of Hermitian positive definite matrices) are respectively named the location and scatter parameters. $\boldsymbol{\tau} \in (\mathbb{R}_+^*)^n$ (set of strictly positive vectors) contains the unknown deterministic texture parameters $\{\tau_i\}_{i=1}^n$. Therefore, $\boldsymbol{x_i}$ follows a Compound Gaussian distribution, i.e

$$\boldsymbol{x_i} \sim \mathbb{CN}(\boldsymbol{\mu}, \tau_i\,\boldsymbol{\Sigma})\,. \qquad (4)$$

One can observe the presence of an indetermination between $\boldsymbol{\tau}$ and $\boldsymbol{\Sigma}$. A determinant constraint $|\boldsymbol{\Sigma}| = 1$ is added to identify the texture and scatter matrix parameters. Hence, $\boldsymbol{\Sigma}$ is defined on the set of unit determinant Hermitian positive definite matrices called $\mathcal{SH}_p^{++}$.

To estimate the three parameters, we want to minimize the negative log-likelihood. The latter is defined on the set $\mathcal{M}_{p,n} = \mathbb{C}^p \times (\mathbb{R}_+^*)^n \times \mathcal{SH}_p^{++}$. $\forall \theta = (\boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\Sigma}) \in \mathcal{M}_{p,n}$, the negative log-likelihood L is written as follow:

$$L(\theta) = \sum_{i=1}^n \left[\log|\tau_i\boldsymbol{\Sigma}| + \frac{(\boldsymbol{x_i}-\boldsymbol{\mu})^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu})}{\tau_i} \right]\,. \quad (5)$$

The MLE θ satisfies the following fixed point equations:

$$\boldsymbol{\mu} = \left(\sum_{i=1}^n \frac{1}{\tau_i}\right)^{-1} \sum_{i=1}^n \frac{\boldsymbol{x_i}}{\tau_i}$$

$$\boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^n \frac{(\boldsymbol{x_i}-\boldsymbol{\mu})(\boldsymbol{x_i}-\boldsymbol{\mu})^H}{\tau_i} \qquad (6)$$

$$\tau_i = \frac{1}{p}\,(\boldsymbol{x_i}-\boldsymbol{\mu})^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu})\,,$$

which coincide with (1) for $u_1(t) = u_2(t) = p/t$, thus naturally extends Tyler's M-estimator to the case of unknown mean. Unfortunately, the fixed-point iterations (2) cannot provide a valid estimation procedure in this context as they do not satisfy Maronna's conditions and generally diverge in practice. An intuitive pathological example can be constructed with $\boldsymbol{\mu}_1^{(k)} = \boldsymbol{x_1}$, that leads $\tau_1^{(k)} = 0$ and causes a division by 0 at the following update of $\boldsymbol{\Sigma}$. Motivated by this issue, we consider the use of Riemannian optimization in order to inherently encode a constraint on the scale of $\boldsymbol{\Sigma}$ and the strict positivity of the variables $\{\tau_i\}_{i=1}^n$. This will result in the formulation of stable algorithms to compute the solution (6).

3. RIEMANNIAN GEOMETRY AND OPTIMIZATION ON $\mathcal{M}_{p,n}$

In this section, we begin by describing the main elements of the manifold of interest $\mathcal{M}_{p,n}$ which embeds the location, texture and scatter matrix parameters and its constraints. The elements of geometry presented are tangent spaces, Riemannian metric, Riemannian exponential mapping and parallel transport. We also introduce the Riemannian distance which is exploited to measure the estimation error. For a full review on this topic, see [12, 13].

Secondly, we present the tools used specifically for estimating location, texture and scatter matrix through Riemannian optimization. Three elements are exhibited: a retraction, a vector transport and the Riemannian gradient of the negative log-likelihood (5).

3.1. Riemannian manifold $\mathcal{M}_{p,n}$

$\mathcal{M}_{p,n}$ is a product manifold of three well known Riemannian manifolds: $\mathbb{C}^p$, $(\mathbb{R}_*^+)^n$ and $\mathcal{SH}_p^{++}$ [14–18]. Thus, the proofs of the following results directly arise from properties of product manifolds [12]. In the following, $\theta = (\boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\Sigma})$, $\theta_1 = (\boldsymbol{\mu}_1, \boldsymbol{\tau}_1, \boldsymbol{\Sigma}_1)$, $\theta_2 = (\boldsymbol{\mu}_2, \boldsymbol{\tau}_2, \boldsymbol{\Sigma}_2)$, $\xi = (\boldsymbol{\xi_\mu}, \boldsymbol{\xi_\tau}, \boldsymbol{\xi_\Sigma})$, $\eta = (\boldsymbol{\eta_\mu}, \boldsymbol{\eta_\tau}, \boldsymbol{\eta_\Sigma})$.

The tangent space of $\mathcal{M}_{p,n}$ at θ denoted $T_\theta\mathcal{M}_{p,n}$ is the product of the tangent spaces of $\mathbb{C}^p$, $(\mathbb{R}_*^+)^n$ and $\mathcal{SH}_p^{++}$ i.e,

$$T_\theta\mathcal{M}_{p,n} = \left\{\xi \in \mathbb{C}^p \times \mathbb{R}^n \times \mathcal{H}_p : \operatorname{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma})=0\right\}\,. \quad (7)$$

$\mathcal{H}_p$ denotes the set Hermitian matrices of size p. Then, we define a Riemannian metric (inner product on tangent spaces), which turns $\mathcal{M}_{p,n}$ into Riemannian manifold. We choose it as the sum of metrics on $\mathbb{C}^p$, $(\mathbb{R}_*^+)^n$ and $\mathcal{SH}_p^{++}$ whose associated geometries are already well known. Let $\xi, \eta \in T_\theta\mathcal{M}_{p,n}$, the Riemannian metric at θ is defined as,

$$\langle\xi,\eta\rangle_\theta^{\mathcal{M}_{p,n}} = \langle\boldsymbol{\xi_\mu},\boldsymbol{\eta_\mu}\rangle_{\boldsymbol{\mu}}^{\mathbb{C}^p} + \langle\boldsymbol{\xi_\tau},\boldsymbol{\eta_\tau}\rangle_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n} + \langle\boldsymbol{\xi_\Sigma},\boldsymbol{\eta_\Sigma}\rangle_{\boldsymbol{\Sigma}}^{\mathcal{H}_p^{++}}\,, \quad (8)$$

with

- $\langle\boldsymbol{\xi_\mu},\boldsymbol{\eta_\mu}\rangle_{\boldsymbol{\mu}}^{\mathbb{C}^p} = \mathfrak{Re}\{\boldsymbol{\xi_\mu}^H\boldsymbol{\eta_\mu}\}$,

- $\langle\boldsymbol{\xi_\tau},\boldsymbol{\eta_\tau}\rangle_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n} = (\boldsymbol{\tau}^{\odot-1}\odot\boldsymbol{\xi_\tau})^T(\boldsymbol{\tau}^{\odot-1}\odot\boldsymbol{\eta_\tau})$, where $\odot$ and $.^{\odot t}$ denote the elementwise product and power operators respectively,

- $\langle\boldsymbol{\xi_\Sigma},\boldsymbol{\eta_\Sigma}\rangle_{\boldsymbol{\Sigma}}^{\mathcal{H}_p^{++}} = \operatorname{Tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta_\Sigma}\right)$.

The corresponding Riemannian exponential mapping at θ (generalization of the concept of straight lines to Riemannian manifolds), denoted $\exp_\theta^{\mathcal{M}_{p,n}}$, is the following application from $T_\theta\mathcal{M}_{p,n}$ onto $\mathcal{M}_{p,n}$,

$$\exp_\theta^{\mathcal{M}_{p,n}}(\xi) = \left(\exp_{\boldsymbol{\mu}}^{\mathbb{C}^p}(\boldsymbol{\xi_\mu}), \exp_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\xi_\tau}), \exp_{\boldsymbol{\Sigma}}^{\mathcal{SH}_p^{++}}(\boldsymbol{\xi_\Sigma})\right), \quad (9)$$

with

- $\exp_{\boldsymbol{\mu}}^{\mathbb{C}^p}(\boldsymbol{\xi_\mu}) = \boldsymbol{\mu} + \boldsymbol{\xi_\mu}$,

- $\exp_{\boldsymbol{\tau}}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\xi_\tau}) = \boldsymbol{\tau} \odot \exp\left(\boldsymbol{\tau}^{\odot -1} \odot \boldsymbol{\xi_\tau}\right)$, where exp is the element-wise exponential operator,

- $\exp_{\boldsymbol{\Sigma}}^{\mathcal{SH}_p^{++}}(\boldsymbol{\xi_\Sigma}) = \boldsymbol{\Sigma} \operatorname{expm}\left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\xi_\Sigma}\right)$, where expm is the matrix exponential.

Then, we present the associated parallel transport between $\theta_1 \in \mathcal{M}_{p,n}$ and $\theta_2 \in \mathcal{M}_{p,n}$, denoted $\mathcal{T}_{\theta_1,\theta_2}^{\mathcal{M}_{p,n}}$. This application moves vectors from the first tangent space $T_{\theta_1}\mathcal{M}_{p,n}$ onto the second one $T_{\theta_2}\mathcal{M}_{p,n}$ while preserving the Riemannian metric. For $\xi \in T_{\theta_1}\mathcal{M}_{p,n}$,

$$\mathcal{T}_{\theta_1,\theta_2}^{\mathcal{M}_{p,n}}(\xi) =$$
$$\left(\mathcal{T}_{\boldsymbol{\mu_1},\boldsymbol{\mu_2}}^{\mathbb{C}^p}(\boldsymbol{\xi_\mu}), \mathcal{T}_{\boldsymbol{\tau_1},\boldsymbol{\tau_2}}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\xi_\tau}), \mathcal{T}_{\boldsymbol{\Sigma_1},\boldsymbol{\Sigma_2}}^{\mathcal{SH}_p^{++}}(\boldsymbol{\xi_\Sigma})\right), \quad (10)$$

with

- $\mathcal{T}_{\boldsymbol{\mu_1},\boldsymbol{\mu_2}}^{\mathbb{C}^p}(\boldsymbol{\xi_\mu}) = \boldsymbol{\xi_\mu}$,

- $\mathcal{T}_{\boldsymbol{\tau_1},\boldsymbol{\tau_2}}^{(\mathbb{R}_*^+)^n}(\boldsymbol{\xi_\tau}) = \boldsymbol{\tau_2} \odot \boldsymbol{\tau_1}^{\odot -1} \odot \boldsymbol{\xi_\tau}$,

- $\mathcal{T}_{\boldsymbol{\Sigma_1},\boldsymbol{\Sigma_2}}^{\mathcal{SH}_p^{++}}(\boldsymbol{\xi_\Sigma}) = \left(\boldsymbol{\Sigma_2}\boldsymbol{\Sigma_1}^{-1}\right)^{\frac{1}{2}} \boldsymbol{\xi_\Sigma} \left(\left(\boldsymbol{\Sigma_2}\boldsymbol{\Sigma_1}^{-1}\right)^{\frac{1}{2}}\right)^H$.

Finally, the distance associated to the Riemannian metric (8) for $\theta_1, \theta_2 \in \mathcal{M}_{p,n}$ is

$$d_{\mathcal{M}_{p,n}}^2(\theta_1, \theta_2) =$$
$$d_{\mathbb{C}^p}^2(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}) + d_{(\mathbb{R}_*^+)^n}^2(\boldsymbol{\tau_1}, \boldsymbol{\tau_2}) + d_{\mathcal{SH}_p^{++}}^2(\boldsymbol{\Sigma_1}, \boldsymbol{\Sigma_2}), \quad (11)$$

with

- $d_{\mathbb{C}^p}^2(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}) = \|\boldsymbol{\mu_1} - \boldsymbol{\mu_2}\|_F^2$,

- $d_{(\mathbb{R}_*^+)^n}^2(\boldsymbol{\tau_1}, \boldsymbol{\tau_2}) = \|\log(\boldsymbol{\tau_1}) - \log(\boldsymbol{\tau_2})\|_F^2$
  where log is the element-wise logarithm operator,

- $d_{\mathcal{SH}_p^{++}}^2(\boldsymbol{\Sigma_1}, \boldsymbol{\Sigma_2}) = \left\|\operatorname{logm}\left(\boldsymbol{\Sigma_1}^{-\frac{1}{2}}\boldsymbol{\Sigma_2}\boldsymbol{\Sigma_1}^{-\frac{1}{2}}\right)\right\|_F^2$
  where logm is the matrix logarithm operator and $\|.\|_F$ is the Frobenius norm.

3.2. Riemannian optimization

As said before, to estimate θ we use the framework [12] to minimize the negative log-likelihood (5). Here, the problem of minimizing L under constraints ($\tau_i > 0$, $\boldsymbol{\Sigma} \succ 0$, and $|\boldsymbol{\Sigma}| = 1$) is reformulated as an unconstrained problem on the Riemannian manifold $\mathcal{M}_{p,n}$ described in the previous subsection, i.e, we search $\hat{\theta} \in \mathcal{M}_{p,n}$ such that

$$\hat{\theta} = \underset{\theta \in \mathcal{M}_{p,n}}{\arg\min} L(\theta). \quad (12)$$

To perform gradient based Riemannian optimization on $\mathcal{M}_{p,n}$, we rely on the three following tools: (*i*) the Riemannian gradient of (5), (*ii*) a retraction and (*iii*) a vector transport. We compute the Riemannian gradient of (5) in Proposition 1. Then, it remains to define a retraction $R_\theta^{\mathcal{M}_{p,n}}$:

$T_\theta \mathcal{M}_{p,n} \to \mathcal{M}_{p,n}$. A natural choice is the exponential mapping (9). However, for numerical stability, we use a second order approximation of exp and expm, i.e, $\exp(\boldsymbol{x}) \approx \boldsymbol{1}_p + \boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^{\odot 2}$ and $\operatorname{expm}(\boldsymbol{X}) \approx \boldsymbol{I}_p + \boldsymbol{X} + \frac{1}{2}\boldsymbol{X}^2$. The vector transport used is the parallel transport given in (10).

Proposition 1 (Riemannian gradient). *The Riemannian gradient L at $\theta \in \mathcal{M}_{p,n}$ is given by*

$$\operatorname{grad} L(\theta) =$$
$$\left(-2\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\frac{\tilde{\boldsymbol{x}_i}}{\tau_i}, p\boldsymbol{\tau} - \mathbf{a}, P_{\boldsymbol{\Sigma}}\left(-\sum_{i=1}^{n}\frac{\tilde{\boldsymbol{x}_i}\tilde{\boldsymbol{x}_i}^H}{\tau_i}\right)\right)$$

where $\tilde{\boldsymbol{x}_i} = \boldsymbol{x}_i - \boldsymbol{\mu}$, $\mathbf{a} \in \mathbb{R}^n$ such that $\mathbf{a}_i = \tilde{\boldsymbol{x}_i}^H\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{x}_i}$. $P_{\boldsymbol{\Sigma}}$ is the orthogonal projection from $\mathcal{H}_p$ onto $T_{\boldsymbol{\Sigma}}\mathcal{SH}_p^{++}$:
$$P_{\boldsymbol{\Sigma}}(\boldsymbol{\xi_\Sigma}) = \boldsymbol{\xi_\Sigma} - \frac{1}{p}\operatorname{Tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}\right)\boldsymbol{\Sigma}. \text{ (See e.g [19]).}$$

Proof. By definition of the Riemannian gradient [12], we have $DL(\theta)[\xi] = \langle \operatorname{grad} L(\theta), \xi \rangle_\theta^{\mathcal{M}_{p,n}}$. Hence, we obtain the directional derivative of L at θ in the direction ξ,

$$DL(\theta)[\xi] = \sum_{i=1}^{n}\left[p\frac{(\boldsymbol{\xi_\tau})_i}{\tau_i} - \frac{(\boldsymbol{\xi_\tau})_i}{\tau_i^2}\mathbf{a}_i - 2\frac{\mathfrak{Re}\{\tilde{\boldsymbol{x}_i}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\mu}\}}{\tau_i}\right.$$
$$\left. - \frac{\tilde{\boldsymbol{x}_i}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{x}_i}}{\tau_i}\right]$$
$$= -2\mathfrak{Re}\left\{\sum_{i=1}^{n}\frac{\tilde{\boldsymbol{x}_i}^H}{\tau_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\mu}\right\}$$
$$+ \left((p\boldsymbol{\tau} - \mathbf{a})\odot\boldsymbol{\tau}^{\odot -1}\right)^T\left(\boldsymbol{\xi_\tau}\odot\boldsymbol{\tau}^{\odot -1}\right)$$
$$- \operatorname{Tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\frac{\tilde{\boldsymbol{x}_i}\tilde{\boldsymbol{x}_i}^H}{\tau_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}\right)$$

To get the Riemannian gradient $\operatorname{grad} L(\theta)$ by identification, it remains to project $\sum_{i=1}^{n}\frac{\tilde{\boldsymbol{x}_i}\tilde{\boldsymbol{x}_i}^H}{\tau_i}$ on $T_{\boldsymbol{\Sigma}}\mathcal{SH}_p^{++}$ with $P_{\boldsymbol{\Sigma}}$. $\square$

With all the tools presented, we optimize (5) with a Riemannian conjugate gradient, briefly summarized in Algorithm 1 (see [12] for more details).

4. NUMERICAL EXPERIMENTS

The purpose of this section is to illustrate the performance of our Riemannian geometry based algorithm.

We estimate location $\boldsymbol{\mu} \in \mathbb{C}^p$ and scatter matrix $\boldsymbol{\Sigma} \in \mathcal{SH}_p^{++}$ from simulated data. Each component of $\boldsymbol{\mu}$ is generated with a complex Normal distribution $\mathbb{CN}(10, 1)$. The scatter matrix matrix is generated as $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$ where $\boldsymbol{U}$ is a random unitary matrix drawn from a complex Normal distribution and $\boldsymbol{\Lambda}$ is a diagonal matrix with each element sampled from a continuous uniform distribution between 1 and 2. Then $\boldsymbol{\Sigma}$ is normalized to have $|\boldsymbol{\Sigma}| = 1$. The texture are randomly generated with positive constraint. In practice, we use

Algorithm 1: Riemannian conjugate gradient [12]

Input : Initial iterate $\theta_1 \in \mathcal{M}_{p,n}$.
Output: Sequence of iterates $\{\theta_k\}$
$k := 1;$
$\xi_1 := -\operatorname{grad} L(\theta_1);$
while *no convergence* **do**
  Compute a step size α_k (e.g see [12, §4.2]) and
    set $\theta_{k+1} := R_{\theta_k}^{\mathcal{M}_{p,n}}(\alpha_k \xi_k);$
  Compute β_{k+1} (e.g see [12, §8.3]) and set
    $\xi_{k+1} := -\operatorname{grad} L(\theta_{k+1}) + \beta_{k+1}\, \mathcal{T}_{\theta_k,\theta_{k+1}}^{\mathcal{M}_{p,n}}(\xi_k);$
 $k := k + 1;$
end

a $\Gamma(0.1, 10)$ distribution to illustrate a non-Gaussian case but other generation processes can be used. The dimension p of data is here equal to 10 and sets $\{\boldsymbol{x_i}\}_{i=1}^n$, with $n \in [\![20, 500]\!]$, are drawn from $\mathbb{CN}(\boldsymbol{\mu}, \tau_i \boldsymbol{\Sigma})$.

The different estimators we compare are the following:

1. Gaussian estimators: sample mean $\boldsymbol{\mu}^{\mathrm{G}}$ and SCM denoted $\boldsymbol{\Sigma}^{\mathrm{G}}$.

2. Two-step estimation: the sets $\{\boldsymbol{x_i}\}_{i=1}^n$ are centered with $\boldsymbol{\mu}^{\mathrm{G}}$ and then we estimate $\boldsymbol{\Sigma}$ using Tyler's M-estimator [11]. The estimator is denoted $\boldsymbol{\Sigma}^{\mathrm{Ty}, \boldsymbol{\mu}^{\mathrm{G}}}$.

3. Tyler's joint estimators of location and scatter matrix [11] denoted $\boldsymbol{\mu}^{\mathrm{Ty}}$ and $\boldsymbol{\Sigma}^{\mathrm{Ty}}$. These estimators corresponds to (1) with $u_1(t) = \sqrt{p/t}$ and $u_2(t) = p/t$. It converges in practice unlike fixed-point equations of the MLE.

4. Tyler's M-estimator with location known [11]. The sets $\{\boldsymbol{x_i}\}_{i=1}^n$ are centered with $\boldsymbol{\mu}$ and then we estimate $\boldsymbol{\Sigma}$. The estimator is denoted $\boldsymbol{\Sigma}^{\mathrm{Ty}, \boldsymbol{\mu}}$.

5. Our estimators $\boldsymbol{\mu}^{\mathrm{CG}}$ and $\boldsymbol{\Sigma}^{\mathrm{CG}}$: a Riemannian conjugate gradient to minimize (5) on $\mathcal{M}_{p,n}$ performed with the library *Pymanopt* [20].

To measure the performance of the considered estimators, we compute the mean squared error (MSE) on location and scatter matrix between the estimators and the real parameters. Figure 1 displays these MSE. The MSE for the estimation of $\boldsymbol{\tau}$ is irrelevant and omitted since the latter vector size grows as the number of samples increases. Our estimators outperform the Gaussian, two-steps and Tyler estimators. This result was expected since our estimators optimize directly the negative log-likelihood (5) which is not the case of Tyler's joint estimators of location and scatter matrix.

Furthermore, we note that for $n \geq 3p$, $\boldsymbol{\Sigma}^{\mathrm{CG}}$ matches the performance of $\boldsymbol{\Sigma}^{\mathrm{Ty}, \boldsymbol{\mu}}$. The latter is known to be globally convergent, consistent, and corresponds to the MLE of the Compound Gaussian for $\boldsymbol{\mu}$ known [11, 21]. This illustrates the good performance of our algorithm that has to estimate both
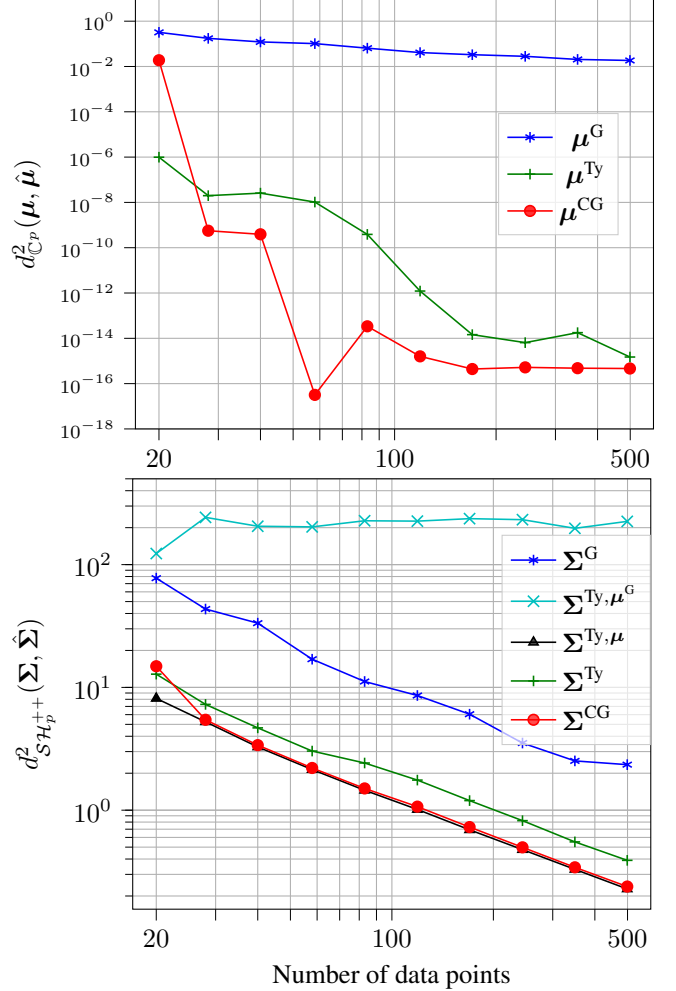
Fig. 1. MSE over 200 simulated sets $\{\boldsymbol{x_i}\}_{i=1}^n$ ($p = 10$) with respect to the number n of samples for the considered estimators $\hat{\boldsymbol{\mu}} \in \{\boldsymbol{\mu}^{\mathrm{G}}, \boldsymbol{\mu}^{\mathrm{Ty}}, \boldsymbol{\mu}^{\mathrm{CG}}\}$ and $\hat{\boldsymbol{\Sigma}} \in \{\boldsymbol{\Sigma}^{\mathrm{G}}, \boldsymbol{\Sigma}^{\mathrm{Ty}, \boldsymbol{\mu}^{\mathrm{G}}}, \boldsymbol{\Sigma}^{\mathrm{Ty}, \boldsymbol{\mu}}, \boldsymbol{\Sigma}^{\mathrm{Ty}}, \boldsymbol{\Sigma}^{\mathrm{CG}}\}$.

location and scatter matrix. For n large enough, the MSE on $\boldsymbol{\mu}^{\mathrm{CG}}$ reaches the machine numerical precision limit and hence becomes negligible, which also illustrates the good performance of our algorithm.

5. CONCLUSIONS

This paper has proposed an efficient Riemannian optimization-based procedure to jointly estimate the location and scatter matrix of a Compound Gaussian distribution. A Riemannian geometry of the parameter manifold $\mathcal{M}_{p,n}$ has been described in order to derive a Riemannian conjugate gradient optimizer. This algorithm reaches performance close to the MLE of the "known location" case, which illustrates the interest of the proposed approach.

6. REFERENCES

[1] D. G. Manolakis, D. Marden, J. P. Kerekes, and G. A. Shaw, "Statistics of hyperspectral imaging data," in *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII*, Sylvia S. Shen and Michael R. Descour, Eds. International Society for Optics and Photonics, 2001, vol. 4381, pp. 308 – 316, SPIE.

[2] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 24–33, 2014.

[3] J. Frontera-Pons, M. A. Veganzones, F. Pascal, and J.-P. Ovarlez, "Hyperspectral anomaly detectors using robust estimators," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 2, pp. 720–731, 2016.

[4] I. S. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1760–1770, 1990.

[5] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Ann. Statist.*, vol. 4, no. 1, pp. 51–67, 01 1976.

[6] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex Elliptically Symmetric distributions: Survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, 2012.

[7] A. Wiesel and T. Zhang, *Structured robust covariance estimation*, Now Foundations and Trends, 2015.

[8] Y. Sun, P. Babu, and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3576–3590, 2016.

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

[10] Y. Sun, P. Babu, and D. P. Palomar, "Regularized robust estimation of mean and covariance matrix under heavy-tailed distributions," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3096–3109, 2015.

[11] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, 03 1987.

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

[13] S. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *Signal Processing, IEEE Transactions on*, vol. 53, pp. 1610 – 1630, 06 2005.

[14] L. T. Skovgaard, "A Riemannian geometry of the multivariate Normal model," *Scandinavian Journal of Statistics*, vol. 11, no. 4, pp. 211–223, 1984.

[15] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *International Journal of computer vision*, vol. 66, no. 1, pp. 41–66, 2006.

[16] R. Bhatia, *Positive Definite Matrices*, Princeton University Press, 2007.

[17] S. Sra and R. Hosseini, "Conic geometric optimization on the manifold of positive definite matrices," *SIAM Journal on Optimization*, vol. 25, no. 1, pp. 713–739, Jan 2015.

[18] F. Bouchard, A. Mian, J. Zhou, S. Said, G. Ginolhac, and Y. Berthoumieu, "Riemannian geometry for compound Gaussian distributions: application to recursive change detection," *Signal Processing*, 2020.

[19] A. Breloy, G. Ginolhac, A. Renaux, and F. Bouchard, "Intrinsic cramér–rao bounds for scatter and shape matrices estimation in CES distributions," *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 262–266, 2019.

[20] J. Townsend, N. Koep, and S. Weichwald, "Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 4755–4759, Jan. 2016.

[21] A. Wiesel, "Geodesic convexity and covariance estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6182–6189, 2012.