Elliptical Wishart Distribution: Maximum Likelihood Estimator from Information Geometry

Imen Ayadi, Florent Bouchard, Frédéric Pascal

ELLIPTICAL WISHART DISTRIBUTION: MAXIMUM LIKELIHOOD ESTIMATOR FROM INFORMATION GEOMETRY

Imen Ayadi, Florent Bouchard, Frédéric Pascal

Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France

ABSTRACT

This work deals with elliptical Wishart distributions on the set of symmetric positive definite matrices. It contains two major contributions. First, the information geometry associated with elliptical Wishart distributions is derived. Second, this geometry is leveraged to propose Riemannian-optimization-based maximum likelihood estimators of any elliptical Wishart distribution. Particular attention is given to two specific distributions: the t- and Kotz Wishart ones. The performance of the proposed methods is assessed through numerical experiments on simulated data.

***Index Terms*—** covariance matrices, robust statistics, elliptical Wishart distribution, information geometry, Riemannian geometry, and optimization

1. INTRODUCTION

Covariance matrices are crucial in various signal processing and machine learning applications, such as radar and image processing [1–3], biomedical signals analysis [4, 5], *etc.*. In these applications, statistics over the set of non-degenerate covariance matrices [6], which is the manifold $\mathcal{S}_p^{++}$ of $p \times p$ symmetric positive definite matrices, are tremendous. Indeed, some statistics over $\mathcal{S}_p^{++}$ are, for instance, leveraged for classification [4, 5] or Bayesian inference [7]. In this context, the most classical distribution on $\mathcal{S}_p^{++}$ is the Wishart distribution [8]. This comes from the fact that it is the distribution of sample covariance matrices of random vectors drawn from a multivariate Gaussian distribution.

However, the Gaussian assumption of the data does not hold in many practical cases. Indeed, in some applications, such as radar processing, data are intrinsically non-Gaussian (see *e.g.,* [9]). In others, due to noise and outliers, it is more likely for data to follow heavy-tailed distributions. In such situations, it is usual to model data with a multivariate elliptical distribution [10]. Analogously to how elliptical distributions generalize the Gaussian one, it is possible to extend the Wishart distribution with the so-called elliptical Wishart distributions [11]. This generalization allows for overcoming some limitations of the Wishart distribution.

Some of the statistical properties of elliptical Wishart distributions are known; see *e.g.*, [12–14]. However, no estimator of the center of the distribution has been derived yet, which is crucial for applications (e.g., classification for electroencephalogram signals). The main contribution of this paper is to propose a maximum likelihood estimator for the center of any elliptical Wishart distribution, with a specific focus on the so-called t- and Kotz-Wishart distributions. To derive these estimators, Riemannian optimization [15] is exploited. The other major contribution of this work is to study the information geometry on $\mathcal{S}_p^{++}$ resulting from elliptical Wishart distributions. Since it is the optimal Riemannian geometry on $\mathcal{S}_p^{++}$ for the considered distributions, it yields the most appropriate Riemannian-optimization-based estimators.

The paper is organized as follows. Section 2 reviews elliptical Wishart distribution while Section 3 analyzes the corresponding information geometry. Then, Section 4 derives maximum likelihood estimators while numerical experiments are presented in Section 5. Finally, concluding remarks and perspectives are drawn in Section 6.

2. ELLIPTICAL WISHART DISTRIBUTION

Elliptical Wishart distributions [11–14] form a large family of distributions on $\mathcal{S}_p^{++}$. They generalize the Wishart distribution in the same way multivariate elliptical distributions generalize the Gaussian one. The probability density function (PDF) associated with the random variable $\boldsymbol{S} \in \mathcal{S}_p^{++}$ following the elliptical Wishart distribution $\mathcal{W}(n, \boldsymbol{\Sigma}, h)$ is, up to a normalization factor,

$$f(\boldsymbol{S}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{S}|^{\frac{n-p-1}{2}} h(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})), \tag{1}$$

where $|\cdot|$ and $\mathrm{tr}(\cdot)$ denote the determinant and trace operators respectively; $h : \mathbb{R}^+ \to \mathbb{R}^+$ is the density generator function of the distribution; $n \geq p$ is an integer corresponding to the degree of freedom; and $\boldsymbol{\Sigma} \in \mathcal{S}_p^{++}$ represents the center of the distribution.

Distributions with PDF (1) can be defined by taking density generator functions of multivariate elliptical distributions

"This work is partially supported by a public grant overseen by the French National Research Agency (ANR) through the program UDOPIA, project funded by the ANR-20-THIA-0013-01"

on $\mathbb{R}^{np}$; see *e.g.*, [10] for a review. Beyond the Gaussian density generator function yielding Wishart, it is possible to define the counterparts of the t-, Kotz, Weibull, *etc.* distributions. While this paper is general, concrete examples are provided for the t-Wishart and Kotz-Wishart distributions. Their density generators functions h are given in Table 1.

As for elliptically contoured vectors, it can be shown that $\boldsymbol{S} \sim \mathcal{W}(n, \boldsymbol{\Sigma}, h)$ admits a stochastic representation, which is essential to obtain the Fisher information metric in Section 3. Due to space limitations, we do not provide proof here. It will be done in a forthcoming paper. It is given by

$$\boldsymbol{S} = \mathcal{Q}\boldsymbol{\Sigma}^{1/2}\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\Sigma}^{1/2}, \tag{2}$$

where $\mathcal{Q} \in \mathbb{R}^+$ and $\boldsymbol{U} \in \mathbb{R}^{p\times n}$ are independent random variables. The PDF of $\mathcal{Q}$ is, up to a factor, $f(\mathcal{Q}) \propto h(\mathcal{Q})\mathcal{Q}^{\frac{np}{2}-1}$. $\boldsymbol{U}$ is such that $\mathrm{vec}(\boldsymbol{U})$ is uniformly distributed on the np-dimensional unit sphere, where $\mathrm{vec}(\cdot)$ denotes the vectorization operator.

Finally, to derive the Fisher information metric and to obtain the maximum likelihood estimator of the elliptical Wishart distribution, it remains to provide the log-likelihood of (1). Given independent and identically distributed (iid) samples $\{\boldsymbol{S}_k\}_{k=1}^K$, the negative log-likelihood corresponding to (1) is, up to an additive constant,

$$\mathcal{L}(\boldsymbol{\Sigma}) = \frac{nK}{2}\log\det(\boldsymbol{\Sigma}) - \sum_{k=1}^{K}\log(h(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_k)). \tag{3}$$

3. INFORMATION GEOMETRY OF THE ELLIPTICAL WISHART DISTRIBUTION

As presented above, the parameter space of the elliptical Wishart distribution is $\mathcal{S}_p^{++}$. Since $\mathcal{S}_p^{++}$ is open in the vector space of symmetric matrices $\mathcal{S}_p$, the tangent space at any $\boldsymbol{\Sigma} \in \mathcal{S}_p^{++}$ can be identified with $\mathcal{S}_p$. The optimal geometry of $\mathcal{S}_p^{++}$ with respect to the elliptical Wishart distribution with negative log-likelihood $\mathcal{L}$ is the one induced by the corresponding Fisher information metric. For $\boldsymbol{\Sigma} \in \mathcal{S}_p^{++}$, $\boldsymbol{\xi}$ and $\boldsymbol{\eta} \in \mathcal{S}_p$, it is given by [16]

$$\langle\boldsymbol{\xi}, \boldsymbol{\eta}\rangle_{\boldsymbol{\Sigma}} = \mathbb{E}[\mathrm{D}^2\,\mathcal{L}(\boldsymbol{\Sigma})[\boldsymbol{\xi}, \boldsymbol{\eta}]]. \tag{4}$$

where $\mathrm{D}^2\,\mathcal{L}(\boldsymbol{\Sigma})[\boldsymbol{\xi}]$ is the second-order directional derivative of $\mathcal{L}$ at $\boldsymbol{\Sigma}$ with respect to the direction $\boldsymbol{\xi}$. When deriving the Fisher metric, it is usual to choose $K = 1$. If needed, the actual Fisher metric for $K \neq 1$ samples is obtained through scaling with K. The Fisher information metric of the elliptical Wishart distribution is provided in Proposition 1. Furthermore, functions Φ and parameters α defined in Proposition 1 are provided in Table 1 for the Wishart, t-Wishart and Kotz-Wishart distributions.

Proposition 1 (Fisher information metric)**.** *The Fisher information metric of the elliptical Wishart distribution with negative log-likelihood* (3) *is, for* $\boldsymbol{\Sigma} \in \mathcal{S}_p^{++}$*,* $\boldsymbol{\xi}$ *and* $\boldsymbol{\eta} \in \mathcal{S}_p$

$$\langle\boldsymbol{\xi}, \boldsymbol{\eta}\rangle_{\boldsymbol{\Sigma}} = \alpha\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}) + \beta\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi})\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}),$$

where, given $\Phi = \frac{h'}{h}$*,*

$$\alpha = \frac{n}{2}\left(1 - \frac{\mathbb{E}[\mathcal{Q}^2\Phi'(\mathcal{Q})]}{\frac{np}{2}(\frac{np}{2}+1)}\right), \quad \beta = \frac{n}{2}\left(\alpha - \frac{n}{2}\right)$$

Proof. Given $K = 1$, the directional derivative of $\mathcal{L}$ at $\boldsymbol{\Sigma}$ with respect to the direction $\boldsymbol{\xi}$ is

$$\mathrm{D}\,\mathcal{L}(\boldsymbol{\Sigma})[\boldsymbol{\xi}] = \frac{n}{2}\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}) + \Phi(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}))\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}).$$

From there, the second-order derivative of $\mathcal{L}$ is

$$\begin{aligned}\mathrm{D}^2\,\mathcal{L}(\boldsymbol{\Sigma})[\boldsymbol{\xi}, \boldsymbol{\eta}] = &-\frac{n}{2}\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta})\\ &- \Phi'(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}))\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi})\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta})\\ &- 2\Phi(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}))\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}).\end{aligned}$$

Finding the Fisher metric thus boils down to computing

$$\begin{aligned}\mathcal{A} &= \mathbb{E}[\Phi(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}))\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta})]\\ \mathcal{B} &= \mathbb{E}[\Phi'(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}))\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi})\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta})].\end{aligned}$$

The calculation of these quantities relies on the stochastic representation $\boldsymbol{S} = \mathcal{Q}\boldsymbol{\Sigma}^{1/2}\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\Sigma}^{1/2}$. One can show that $\mathbb{E}[\boldsymbol{U}\boldsymbol{U}^T] = \frac{1}{p}\boldsymbol{I}_{\boldsymbol{p}}$ and $\boldsymbol{V}_{\boldsymbol{U}} = \mathbb{E}[\mathrm{vec}(\boldsymbol{U}\boldsymbol{U}^T)\,\mathrm{vec}(\boldsymbol{U}\boldsymbol{U}^T)^T] = (\boldsymbol{I}_{p^2} + \boldsymbol{K}_{pp} + n\,\mathrm{vec}(\boldsymbol{I}_p)\,\mathrm{vec}(\boldsymbol{I}_p)^T)/(p(np+2))$, where $\boldsymbol{K}_{pp}$ denotes the commutation matrix. Since $\mathbb{E}[\mathcal{Q}\Phi(\mathcal{Q})] = -\frac{np}{2}$,

$$\mathcal{A} = \mathbb{E}[\mathcal{Q}\Phi(\mathcal{Q})]\,\mathrm{tr}(\mathbb{E}[\boldsymbol{U}\boldsymbol{U}^T]\bar{\boldsymbol{\xi}}\bar{\boldsymbol{\eta}}) = -\frac{n}{2}\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}),$$

where $\bar{\boldsymbol{\xi}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-1/2}$ and $\bar{\boldsymbol{\eta}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}\boldsymbol{\Sigma}^{-1/2}$. Furthermore, since $\mathrm{tr}(\boldsymbol{X}^T\boldsymbol{Y}) = \mathrm{vec}(\boldsymbol{X})^T\,\mathrm{vec}(\boldsymbol{Y})$, one can show

$$\begin{aligned}\mathcal{B} &= \mathbb{E}[\mathcal{Q}^2\Phi'(\mathcal{Q})]\,\mathrm{vec}(\bar{\boldsymbol{\xi}})^T\boldsymbol{V}_{\boldsymbol{U}}\,\mathrm{vec}(\bar{\boldsymbol{\eta}}),\\ &= \tfrac{\mathbb{E}[\mathcal{Q}^2\Phi'(\mathcal{Q})]}{p(np+2)}(2\,\mathrm{tr}(\bar{\boldsymbol{\xi}}\bar{\boldsymbol{\eta}}) + n\,\mathrm{tr}(\bar{\boldsymbol{\xi}})\,\mathrm{tr}(\bar{\boldsymbol{\eta}})).\end{aligned}$$

Basic manipulations conclude the proof. ∎

From the Cauchy-Schwarz inequality, the Fisher metric of Proposition 1 defines a proper Riemannian metric only if $\alpha > 0$ and $\alpha + p\beta > 0$. With an integration by parts, one can show $\mathbb{E}[\mathcal{Q}^2\Phi'(\mathcal{Q})] = \frac{np}{2}\left(\frac{np}{2}+1\right) - \mathbb{E}[\mathcal{Q}^2\Phi(\mathcal{Q})^2]$. Thus, the condition is equivalent to $\mathrm{var}[\mathcal{Q}\Phi(\mathcal{Q})] > 0$, which is fulfilled as long as $\mathcal{Q}\Phi(\mathcal{Q})$ is not a constant almost surely. This happens to be true for every elliptical distribution.

Interestingly, the Fisher information metric of elliptical Wishart distribution shares the form of the Fisher information metric of a multivariate elliptical distribution (with different values of α and β) [17]. It is expected because one can show that elliptical Wishart distributions are closely linked to multivariate elliptical distributions with a Kronecker product structured covariance matrix, whose information geometry is derived in [18].

The geometry of $\mathcal{S}_p^{++}$ equipped with a Riemannian metric of the form given in Proposition 1 is well-known and can be

Distribution	Wishart	Student Wishart	Kotz Wishart
$h(t)$	$\exp(-t/2)$	$(\nu+t)^{-\frac{\nu+np}{2}}, \quad \nu > 0$	$t^{a-1}\exp(-rt^b), \quad a > 1 - \frac{np}{2}$; b, $r > 0$
$\Phi(t)$	$-\frac{1}{2}$	$-\frac{1}{2}\frac{\nu+np}{\nu+t}$	$-rbt^{b-1} + \frac{a-1}{t}$
α	$\frac{n}{2}$	$\frac{n}{2}\frac{\nu+np}{\nu+np+2}$	$\frac{np(npb+2)+b(1-a)}{2pb(np+2)}$

Table 1. Density generator functions h, functions Φ and Fisher metric parameter α of the Wishart, t-Wishart (with ν degrees of freedom) and Kotz Wishart (with parameters a, b and r) distributions.

found, for instance, in [17]. In particular, the geodesic (a generalization of a straight line on a manifold) $\gamma : [0,1] \to \mathcal{S}_p^{++}$ emanating from $\boldsymbol{\Sigma} \in \mathcal{S}_p^{++}$ in the direction $\boldsymbol{\xi} \in \mathcal{S}_p$ is

$$\gamma(t) = \boldsymbol{\Sigma}\exp(t\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}), \tag{5}$$

where $\exp(\cdot)$ denotes the matrix exponential. From there, one can deduce the Fisher information distance of the elliptical Wishart distribution. It is defined as the length of the geodesic joining $\boldsymbol{\Sigma}$ and $\overline{\boldsymbol{\Sigma}}$ according to the metric of Proposition 1 and its square is equal to

$$\delta^2(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) = \alpha\left\|\log(\boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2})\right\|_2^2 + \beta\left(\log\det(\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}})\right)^2. \tag{6}$$

where $\|\cdot\|_2$ is the Frobenius norm and $\log(\cdot)$ denotes the matrix logarithm.

4. MAXIMUM LIKELIHOOD ESTIMATOR FROM RIEMANNIAN OPTIMIZATION

The maximum likelihood estimator is the solution of the constrained optimization problem

$$\underset{\boldsymbol{\Sigma}\in\mathcal{S}_p^{++}}{\operatorname{argmin}} \quad \mathcal{L}(\boldsymbol{\Sigma}). \tag{7}$$

For the Wishart distribution, the solution is known in closed form. Indeed, it is simply

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{W}} = \frac{1}{nK}\sum_k \boldsymbol{S}_k. \tag{8}$$

However, as for multivariate elliptical distributions, no closed-form solution is known in the general case, and one needs an iterative algorithm to find the solution. In this work, we leverage the geometry of $\mathcal{S}_p^{++}$ provided in Section 3 to solve (7) through Riemannian optimization [15].

To do so, the first step to define is the Riemannian gradient $\nabla_{\mathcal{S}_p^{++}}\mathcal{L}(\boldsymbol{\Sigma})$ of $\mathcal{L}$ at $\boldsymbol{\Sigma} \in \mathcal{S}_p^{++}$. It is defined as the only tangent vector such that, for all $\boldsymbol{\xi} \in \mathcal{S}_p$,

$$\langle \nabla_{\mathcal{S}_p^{++}}\mathcal{L}(\boldsymbol{\Sigma}), \boldsymbol{\xi}\rangle_{\boldsymbol{\Sigma}} = \mathrm{D}\,\mathcal{L}(\boldsymbol{\Sigma})[\boldsymbol{\xi}].$$

Rather than computing the Riemannian gradient directly, it is often handy to calculate the Euclidean gradient $\nabla_{\mathcal{E}}\mathcal{L}$ and then deduce the Riemannian one. From [19], in this case, one has

$$\nabla_{\mathcal{S}_p^{++}}\mathcal{L}(\boldsymbol{\Sigma}) = \frac{1}{\alpha}\boldsymbol{\Sigma}\nabla_{\mathcal{E}}\mathcal{L}(\boldsymbol{\Sigma})\boldsymbol{\Sigma} - \frac{\beta}{\alpha(\alpha+p\beta)}\operatorname{tr}(\nabla_{\mathcal{E}}\mathcal{L}(\boldsymbol{\Sigma})\boldsymbol{\Sigma})\boldsymbol{\Sigma}.$$

The Euclidean gradient $\nabla_{\mathcal{E}}\mathcal{L}$ of the negative log-likelihood of an elliptical Wishart distribution is given in Proposition 2. Functions Φ required to compute the Euclidean gradient in practice are provided in Table 1 for the Wishart, t-Wishart and Kotz Wishart distributions.

Proposition 2 (Euclidean gradient)**.** *The Euclidean gradient of the negative log-likelihood $\mathcal{L} : \mathcal{S}_p^{++} \to \mathbb{R}$ defined in* (3) *of the elliptical Wishart distribution is, for all $\boldsymbol{\Sigma} \in \mathcal{S}_p^{++}$,*

$$\nabla_{\mathcal{E}}\mathcal{L}(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1}\left(\frac{nK}{2}\boldsymbol{\Sigma} + \sum_{k=1}^{K}\Phi(\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_k))\boldsymbol{S}_k\right)\boldsymbol{\Sigma}^{-1},$$

Proof. This results directly from $\nabla_{\mathcal{E}}\log\det(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1}$ and $\nabla_{\mathcal{E}}tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}) = -\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}$. ■

The Riemannian gradient is sufficient to define a descent direction of $\mathcal{L}$ at $\boldsymbol{\Sigma}$, yielding the Riemannian steepest descent algorithm [15]. However, if one wants to employ more sophisticated optimization methods such as conjugate gradient or the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, defining a vector transport operator. Such methods allow transporting a tangent vector of one point onto the tangent space at another point, [15] is needed. The most natural choice is the one corresponding to parallel transport on $\mathcal{S}_p^{++}$ [20]. The transport of tangent vector $\boldsymbol{\xi}$ of $\boldsymbol{\Sigma}$ onto the tangent space at $\widehat{\boldsymbol{\Sigma}}$ is $\mathcal{T}_{\boldsymbol{\Sigma}\to\widehat{\boldsymbol{\Sigma}}}(\boldsymbol{\xi}) = (\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1})^{1/2}\boldsymbol{\xi}(\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}})^{1/2}$.

Once a descent direction is selected, it remains to get from the tangent space back onto the manifold. This is achieved by a retraction [15]. In our case, from a numerical perspective, the best solution is to take the second-order approximation of the geodesics (5). Given $\boldsymbol{\Sigma} \in \mathcal{S}_p^{++}$ and $\boldsymbol{\xi} \in \mathcal{S}_p$, it is [20]

$$R_{\boldsymbol{\Sigma}}(\boldsymbol{\xi}) = \boldsymbol{\Sigma} + \boldsymbol{\xi} + \frac{1}{2}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}.$$

With these tools, a large panel of Riemannian optimization algorithms can be employed to solve (7). For instance, the sequence of iterates $\{\boldsymbol{\Sigma}_i\}$ and descent directions $\{\boldsymbol{\xi}_i\}$ generated by a Riemannian conjugate gradient algorithm is

$$\begin{aligned}\boldsymbol{\Sigma}_{i+1} &= R_{\boldsymbol{\Sigma}_i}(\boldsymbol{\xi}_i)\\ \boldsymbol{\xi}_i &= t_i(-\nabla_{\mathcal{S}_p^{++}}\mathcal{L}(\boldsymbol{\Sigma}_i) + \beta_i\mathcal{T}_{\boldsymbol{\Sigma}_{i-1}\to\boldsymbol{\Sigma}_i}(\boldsymbol{\xi}_{i-1})),\end{aligned} \tag{9}$$

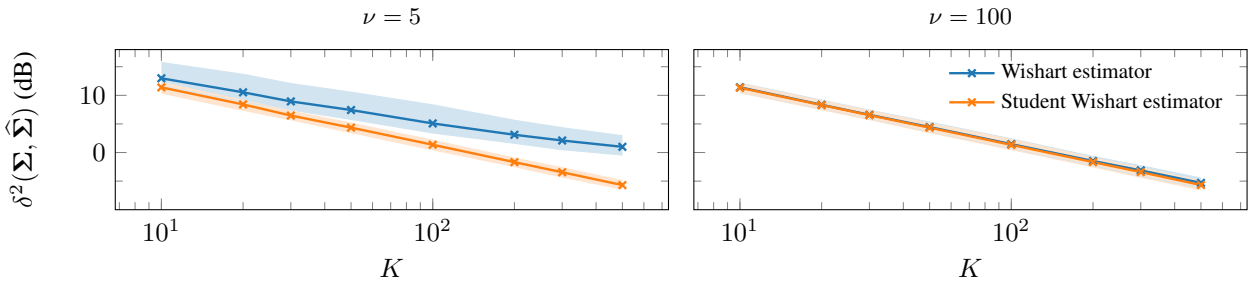

Fig. 1. Median (dark lines), 5% and 95% quantiles (filled areas) of error measures of maximum likelihood estimators of Wishart and Student Wishart distributions as functions of the number of samples K. Medians and quantiles are computed over 200 simulated sets $\{\boldsymbol{S}_k\}$ drawn from a Student Wishart distribution.

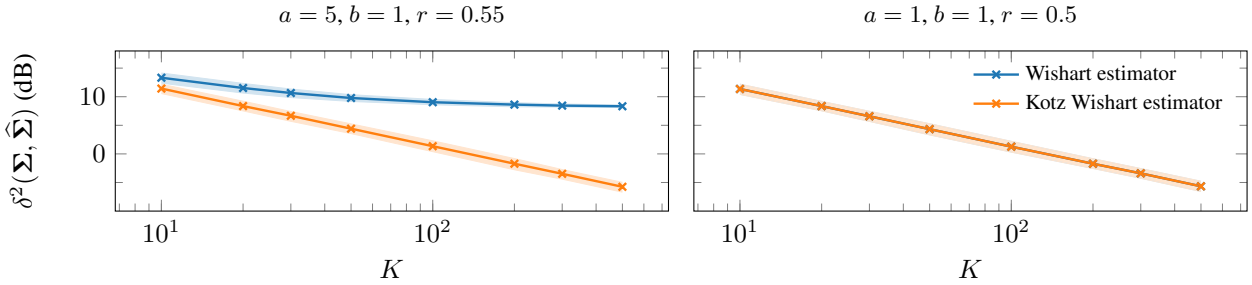

Fig. 2. Median (dark lines), 5% and 95% quantiles (filled areas) of error measures of maximum likelihood estimators of Wishart and Kotz Wishart distributions as functions of the number of samples K. Medians and quantiles are computed over 200 simulated sets $\{\boldsymbol{S}_k\}$ drawn from a Kotz Wishart distribution.

where t_i is a stepsize computed through a linesearch [15] and β_i can be computed using the rule in [21] for example.

5. NUMERICAL EXPERIMENTS

To validate the interest of the maximum likelihood estimators derived in Section 4, numerical experiments are conducted on simulated data drawn from the t- and Kotz-Wishart distributions. We set $n = 100$, $p = 16$ and we randomly generate a center $\boldsymbol{\Sigma} \in \mathcal{S}_p^{++}$. For $K \in \{30, 70, 100, 300, 500\}$, we draw iid samples $\{\boldsymbol{S}_k\}_{k=1}^K$ according to $\mathcal{W}(n, \Sigma, h)$, where h corresponds either to the t-Wishart distribution with $\nu \in \{5, 100\}$ or to the Kotz-Wishart distribution with $(a, b, r) \in \{(1, 1, 0.5), (5, 1, 0.55)\}$. For each setting, 200 data sets are simulated.

For each kind of simulated data (t- or Kotz-Wishart), we compare two different estimation algorithms: the Wishart estimator $\widehat{\boldsymbol{\Sigma}}_{\text{W}}$ defined in (8); and the maximum likelihood estimator corresponding to the simulated data (t- or Kotz Wishart). The latter are computed with a Riemannian conjugate gradient algorithm as presented in (9). To evaluate the estimation error, we employ the Fisher information distance (6) between the true center and estimators.

Figures 1 and 2 display the medians, 5%, and 95% quantiles of the errors of each considered estimator. When data follow (or are close to follow) the Wishart distribution, *i.e.*, t-Wishart with $\nu = 100$ or Kotz-Wishart with $(a, b, r) = (1, 1, 0.5)$, we observe that the maximum likelihood estimator of the true distribution obtained via Riemannian optimization and the Wishart estimator feature equivalent performance. In such cases, the Wishart estimator is preferred as it is much cheaper to compute. However, as expected, when we stray away from the Wishart distribution, *i.e.*, t-Wishart with $\nu = 5$ or Kotz-Wishart with $(a, b, r) = (5, 1, 0.55)$, the Wishart estimator no longer provides good results and is strongly outperformed by maximum likelihood estimators.

6. CONCLUSION AND PERSPECTIVES

This paper studies elliptical Wishart distributions, providing two significant contributions. The first consists in deriving the information geometry on $\mathcal{S}_p^{++}$ associated with this family of distributions. The second is to develop a Riemannian-optimization-based method to compute the maximum likelihood estimator of elliptical Wishart distributions. The excellent performance of the proposed methods is validated through numerical experiments.

In future works, the geodesic convexity of the negative log-likelihood (3) will be studied to prove the convergence of the proposed algorithm adequately. Moreover, the optimal reachable performance of the estimators will be investigated by finding the intrinsic Cramér-Rao lower bound [16]. Finally, resolving the estimation problem will enable to enlarge the use of elliptical Wishart distributions in signal processing.

7. REFERENCES

[1] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1713–1727, 2008.

[2] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 34–48, 2007.

[3] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu, "Parameter estimation for multivariate generalized Gaussian distributions," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5960–5971, 2013.

[4] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2011.

[5] S. Chevallier, E. K. Kalunga, Q. Barthélemy, and E. Monacelli, "Review of Riemannian distances and divergences, applied to SSVEP-based BCI," *Neuroinformatics*, vol. 19, no. 1, pp. 93–106, 2021.

[6] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*, Chapman and Hall/CRC, 2018.

[7] O. Besson, S. Bidon, and J.-Y. Tourneret, "Covariance matrix estimation with heterogeneous samples," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 909–920, 2008.

[8] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20A, no. 1-2, pp. 32–52, 12 1928.

[9] K. J. Sangston and K. R/ Gerlach, "Coherent detection of radar targets in a non-Gaussian background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 2, pp. 330–340, 1994.

[10] 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.

[11] K.-T. Fang and T. W. Anderson, *Statistical inference in elliptically contoured and related distributions*, Allerton Press, 1990.

[12] F. J. Caro-Lopera, G. González-Farías, and N. Balakrishnan, "On generalized Wishart distributions-I: Likelihood ratio test for homogeneity of covariance matrices," *Sankhya A*, vol. 76, no. 2, pp. 179–194, 2014.

[13] A. Bekker, M. Arashi, and J. Van Niekerk, "Wishart generator distribution," *arXiv preprint arXiv:1502.07300*, 2015.

[14] F. J. Caro-Lopera, G. González-Farías, and N. Balakrishnan, "Matrix-variate distribution theory under elliptical models-4: Joint distribution of latent roots of covariance matrix and the largest and smallest latent roots," *Journal of Multivariate Analysis*, vol. 145, pp. 224–235, 2016.

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

[16] S. T. Smith, "Covariance, subspace, and intrinsic Cramér–Rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, 2005.

[17] 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, 2018.

[18] F. Bouchard, A. Breloy, A. Mian, and G. Ginolhac, "Online Kronecker product structured covariance estimation with Riemannian geometry for t-distributed data," in *2021 29th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 856–859.

[19] F. Bouchard, A. Breloy, G. Ginolhac, A. Renaux, and F. Pascal, "A Riemannian framework for low-rank structured elliptical models," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1185–1199, 2021.

[20] B. Jeuris, R. Vandebril, and B. Vandereycken, "A survey and comparison of contemporary algorithms for computing the matrix geometric mean," *Electronic Transactions on Numerical Analysis*, vol. 39, pp. 379–402, 2012.

[21] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear equation," *Journal of research of the National Bureau of Standards*, vol. 49, no. 6, pp. 409, 1952.