Intrinsic Cramér-Rao Bounds for Scatter and Shape Matrices Estimation in Complex Elliptically Symmetric Distributions

Arnaud Breloy, Guillaume Ginolhac, Alexandre Renaux, Florent Bouchard

December 4, 2018

Abstract

Scatter matrix and its normalized counterpart, referred to as shape matrix, are key parameters in multivariate statistical signal processing, as they generalize the concept of covariance matrix in the widely used Complex Elliptically Symmetric distributions. Following the framework of [1], intrinsic Cramér-Rao bounds are derived for the problem of scatter and shape matrices estimation with samples following a Complex Elliptically Symmetric distribution. The Fisher Information Metric and its associated Riemannian distance (namely, CES-Fisher) on the manifold of Hermitian positive definite matrices are derived. Based on these results, intrinsic Cramér-Rao bounds on the considered problems are then expressed for three different distances (Euclidean, natural Riemannian and CES-Fisher). These contributions are therefore a generalization of Theorems 4 and 5 of [1] to a wider class of distributions and metrics for both scatter and shape matrices.

Performance Analysis, Intrinsic Cramér-Rao, Fisher information, Riemannian geometry, CES distributions, covariance, scatter, Shape, M-estimators.

0.1 Introduction

Cramér-Rao lower bounds are ubiquitous tools in statistical signal processing, as they characterize the optimum performances in terms of mean squared error that can be achieved for a given parametric estimation problem [2]. Hence they are usually used to assess the performance of an estimation process, but they can also provide a criterion to optimize the parameters of a system. In some contexts, the parameters to be estimated are inherently satisfying a system of constraints (e.g. positiveness, normalization...), which, once taken into account in the estimation process, translates in gain in estimation accuracy. To reflect this gain, that does not appear in the standard analysis, the so-called constrained Cramér-Rao bounds have been developed in [3–5]. However, for parameters living in a manifold (e.g. positive definite matrices, subspaces, rotation matrices, ...) the constraints often may not be explicited in a simple system of equations, as there is no intrinsic set of coordinates. For example, a general expression of the Cramér-Rao Bounds for the estimation of the covariance matrix using a parameterization $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$ with appropriate parameters vector $\boldsymbol{\theta}$ (typically, the real and imaginary parts of the covariance matrix entries) is given in the form of the inequality

$$\mathbb{E}\left[\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|_F^2\right] \geq \operatorname{Tr}\left\{\mathbf{F}^{-1}\right\}, \tag{1}$$

where $\mathbf{F}$ is the Fisher information matrix. However, notice that $\boldsymbol{\theta}$ is treated here as a vector with arbitrary values, which does not ensure that $\mathbf{\Sigma}(\boldsymbol{\theta})$ is positive definite. There are also no practical explicit constraints formulations on a parameterization $\mathbf{\Sigma}(\boldsymbol{\theta})$ to ensure this property. Therefore, a constrained Cramér-Rao bound [3–5] cannot be derived for a more refined or meaningful performance study. Additionally, the classical Cramér-Rao analysis provides a lower bound on the mean squared error (Euclidean metric), while this criterion may not be the most appropriate for characterizing the performance in a given context. Especially, for parameters living in a manifold, it can be more relevant to caracterize a lower bound on the mean natural Riemannian distance between the true parameters and the estimators, which can also reveal hidden properties of estimators.

To overcome these two issues, intrinsic (i.e. in a manifold setting) versions of the Cramér-Rao inequality have been established in [1, 6–10], and, for example, applied to rotation matrices estimation problems in [11, 12]. Notably, in [1] intrinsic bounds are expressed in the form of a matrix inequality between the covariance of the inverse exponential map and the Fisher information matrix, which is valid for any chosen Riemannian metric. Hence, this inequality allows to obtain Cramér-Rao bounds for various distances (depending on the chosen metric). In [1], these bounds are then applied to the problems of covariance matrix and subspace estimation for samples following a multivariate Gaussian distribution (for both real and complex case). Covariance matrix estimation is a fundamental issue in signal processing and this intrinsic analysis offered a lower bound on a relevant performance criterion (the natural Riemannian distance between Hermitian positive definite matrices) as well as interesting insights, e.g. the observation of a bias at low sample support that is not exhibited by the traditional Cramér-Rao analysis.

The aim of this work is to generalize this intrinsic analysis to the class of Complex Elliptically Symmetric (CES) distributions [13–15]. The multivariate CES distributions are defined by a density generator, a center of distribution, and a scatter matrix (referred to as shape matrix if normalized), which is proportional to the covariance matrix if the latter exists. These distributions provide a class that has recently attracted interest in the signal processing community, as it includes a large panel of well known distributions such as Weibull, Student t-distribution, Generalized Gaussian, K-distribution... that can accurately model various physical phenomenon, such as radar clutter measurements [16, 17] or observations in image processing [18–20]. Additionally, CES distributions share connections with M-estimators and the robust estimation theory [21]. This general framework has been extensively used in the modern estimation/detection literature due to its interest, both from a theoretical and practical point of view (see e.g. [15] and references therein). The concepts and tools from Riemannian geometry also raised recent interest in this context, as they allow to reveal hidden convexity of likelihood functions, as well as designing practical regularization penalties for M-estimators [22–25]. Non-intrinsic Cramér-Rao bounds for scatter matrix estimation have been established for several CES distribution in [26–28], and a generalized Slepian-Bangs formula to obtain the entries of the Fisher information matrix for CES distribution is derived in [29]. However, these studies are performed in the traditional Euclidean setting and do not offer the aforementioned advantages of an intrinsic analysis.

Using the framework of [1], we derive in this paper intrinsic Cramér-Rao bounds for the problem of scatter and shape matrix estimation with samples following a CES distribution [15]. To this aim, we first study the information geometry induced by a CES likelihood on the set of Hermitian positive definite matrices. We derive the Fisher information metric and its associated Riemannian distance (namely, CES-Fisher). As a byproduct, we fully describe the Riemannian geometry of the positive definite matrices manifold equipped with this metric (Levi-Civita connection, Geodesics). We note that this CES-Fisher metric have also been studied in [30, 31] but these results are not developed for data with complex entries, and our derivations propose an alternate, more concise proof. Second, based on the previous results, we derive intrinsic Cramér-Rao bounds on the considered problem for three different distances (Euclidean, natural Riemannian and CES-Fisher) for both the scatter and shape matrix parameters. These contributions provide therefore a generalization of Theorems 4 and 5 of [1] to a wider class of distribution and metrics for the scatter matrix. The Cramér-Rao bounds derived on the shape matrix are original for both Euclidean and Riemannian distances. Notably, these results allow to draw a fair comparison of different M-estimators, regardless of the scaling ambiguities inherent to CES distributions.

The paper is organized as follows. Section II-A and II-B provide an introduction to CES distributions, which is oriented on the problem of scatter matrix estimation in this context. We also emphasize on a scaling ambiguity naturally brought by CES distributions, leading to the definition of shape matrix. Section II-C presents the background on intrinsic Cramér-Rao bounds. Section III deals with the information geometry induced by a CES likelihood: we derive the Fisher information metric and its associated Riemannian distance (namely CES-Fisher) on the set of Hermitian positive definite matrices. Section IV builds upon these results to derive intrinsic Cramér-Rao bounds on the problem of scatter matrix estimation in CES distributions for different distances: Euclidean distance, natural Riemannian distance and CES-Fisher distance. Section V extends these results on the shape matrix estimation problem. Section VI illustrates these results with Monte-Carlo simulations for the multivariate Student t-distribution.

The following convention is adopted: italic indicates a scalar quantity, lower case boldface indicates a vector quantity and upper case boldface a matrix. $\cdot^H$ denotes the transpose conjugate operator or the simple conjugate operator for a scalar quantity. For a function of a real parameter g, $g'(t) = \mathrm{d}g(t)/\mathrm{d}t$ denotes its derivative. $\mathcal{H}_M$ is the manifold of $M \times M$ hermitian matrices. $\mathcal{H}_M^+$ is the manifold of $M \times M$ hermitian nonnegative definite matrices. $\mathcal{H}_M^{++}$ is the manifold of $M \times M$ hermitian positive definite matrices. $\mathcal{SH}_M^{++}$ is the special group of $\mathcal{H}_M^{++}$, i.e. the the manifold of $M \times M$ Hermitian positive definite matrices with unit determinant. $\mathbf{A} \succcurlyeq \mathbf{B}$ means $\mathbf{A} - \mathbf{B} \in \mathcal{H}_M^+$ for $\mathbf{A}, \mathbf{B} \in \mathcal{H}_M$. $\mathrm{D}\,\mathbf{\Omega}_2[\mathbf{\Omega}_1]$ denotes the directional derivative of $\mathbf{\Omega}_2$ in the direction $\mathbf{\Omega}_1$. $\mathcal{CES}(\mathbf{a}, \mathbf{\Sigma}, g)$ is a complex elliptic symmetric vector of mean $\mathbf{a}$, scatter matrix $\mathbf{\Sigma}$ and density generator g. $\mathbb{E}\,[\,]$ is the expectation operator. $\mathbf{I}_M$ is the $M \times M$ identity matrix. $|\mathbf{\Sigma}|$ is the determinant of the matrix $\mathbf{\Sigma}$ and $\mathrm{Tr}\{\}$ is the Trace operator. $\hat{\boldsymbol{\theta}}$ is an estimate of the parameter $\boldsymbol{\theta}$. $\{w_n\}_{n \in [\![1,N]\!]}$ denotes the set of n elements w_n with $n \in [\![1,N]\!]$ and whose writing will often be contracted into $\{w_n\}$. $\mathrm{diag}(a_n)$ is the $N \times N$ diagonal matrix with diagonal elements a_n. $\mathbf{0}_{M \times N}$ (respectively $\mathbf{1}_{M \times N}$) denotes the $M \times N$ matrices with zeros (respectively ones) in all entries. $\delta_{i,j}$ denotes the Kronecker delta applied to the couple (i, j).

0.2 Background

0.2.1 CES distributions, Scatter and Shape matrices

Complex Elliptically Symmetric (CES) distributions [13] refer to a large family of multivariate distributions. We refer the reader to the very comprehensive and detailed review on the topic in the references [14, 15], from which we adopt most of the formalism. A vector $\mathbf{z} \in \mathbb{C}^M$ follows a zero-mean CES distribution, denoted $\mathbf{z} \sim \mathcal{CES}\left(\mathbf{0}, \mathbf{\Sigma}, g\right)$, if it admits the following stochastic representation:

$$\mathbf{z} \stackrel{d}{=} \sqrt{\mathcal{Q}}\, \mathbf{\Sigma}^{1/2}\, \mathbf{u}, \tag{2}$$

where:
- The notation $\stackrel{d}{=}$ stands for "has the same distribution as".
- The vector $\mathbf{u} \in \mathbb{C}^M$ follows a uniform distribution on the complex unit sphere $\mathbb{C}\mathcal{H}^M = \left\{\mathbf{u} \in \mathbb{C}^M \mid \|\mathbf{u}\| = 1\right\}$, denoted $\mathbf{u} \sim \mathcal{U}\left(\mathbb{C}S^M\right)$.
- The scalar $\mathcal{Q} \in \mathbf{R}_+$ is non-negative real random variable of probability density function p, independent of $\mathbf{u}$, and called the second-order modular variate (while $\sqrt{\mathcal{Q}}$ is called the modular variate).
- The matrix $\mathbf{\Sigma}^{1/2} \in \mathbb{C}^{M \times M}$ is a factorization of the scatter matrix $\mathbf{\Sigma} = \mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{H/2}$. If

the covariance matrix of $\mathbf{z}$ exists, it is proportional to the scatter matrix, i.e. $\mathbb{E}\left[\mathbf{z}\mathbf{z}^H\right] \propto \boldsymbol{\Sigma}$.

We focus here only on the absolute-continuous case, i.e. when the scatter matrix $\boldsymbol{\Sigma}$ is full rank. In this case, the probability density function of $\mathbf{z}$ is given as

$$f\left(\mathbf{z}|\boldsymbol{\Sigma}, g\right) \propto |\boldsymbol{\Sigma}|^{-1} g\left(\, \mathbf{z}^H \boldsymbol{\Sigma}^{-1} \mathbf{z}\,\right), \tag{3}$$

where the function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is called the density generator. The density generator is satisfying the finite moment condition $\delta_{M,g} = \int_0^\infty t^{M-1} g(t) dt < \infty$. This function g is related to the probability density function of the second-order modular variate by the relation

$$p\left(\mathcal{Q}\right) = \delta_{M,g}^{-1} \mathcal{Q}^{M-1} g\left(\mathcal{Q}\right). \tag{4}$$

Notice that the definition of a CES distribution naturally presents a scaling ambiguity. Indeed, consider $\tau \in \mathbb{R}_+^*$, the couples $\{\mathcal{Q},\ \boldsymbol{\Sigma}\}$ and $\{\mathcal{Q}/\tau,\ \tau\boldsymbol{\Sigma}\}$ lead to the same distribution of $\mathbf{z}$ according to (2). This ambiguity is not impactful, as most of adaptive processes only require an estimate of the scatter matrix up to a scale [17]. To this end, let us define

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{V}, \tag{5}$$

where $\mathbf{V}$ denotes the normalized scatter matrix, called the shape matrix, and the scalar σ^2 is referred to as scale parameter. In the following we will chose the canonical unitary determinant normalization advocated in [32]. Indeed, we will show that this choice appears natural w.r.t. the geometry of the problem and offers a practical and meaningful view for intrinsic performance analysis. Hence $\mathbf{V}$ belongs to the manifold referred to as the special group of $\mathcal{H}_M^{++}$, denoted

$$\mathcal{SH}_M^{++} = \left\{\mathbf{V} \in \mathcal{H}_M^{++},\ |\mathbf{V}| = 1\right\}. \tag{6}$$

Eventually, a simple way to redefine a CES distribution $\mathbf{z} \sim \mathcal{CES}\left(\mathbf{0}, \boldsymbol{\Sigma}, g\right)$ so that scale and shape parameters coincide is to absorb the ambiguity in the second-order modular variate as $\mathcal{Q}' \stackrel{d}{=} \sqrt[M]{|\boldsymbol{\Sigma}|}\mathcal{Q}$, leading to the equivalent distribution $\mathbf{z} \sim \mathcal{CES}\left(\mathbf{0}, \mathbf{V}, \tilde{g}\right)$, where $\tilde{g}$ is appropriately set from $p(\mathcal{Q}')$ and the relation in (4). Also note that some other normalization of the shape exist [32], such as the constrained trace used in [15]. However, these alternate normalizations often define manifolds with unknown geodesics, which is not suited for the present intrinsic analysis.

0.2.2 Scatter and Shape estimation in CES distributions

In this section, we focus on the ubiquitous problem of estimating the scatter and shape matrices from a sample set $\{\mathbf{z}_k\}_{k\in[\![1,K]\!]}$ (with shortened notation $\{\mathbf{z}_k\}$), distributed as $\mathbf{z} \sim \mathcal{CES}\left(\mathbf{0}, \boldsymbol{\Sigma}, g\right)$ [15, 33–35]. Assuming a zero-mean CES distribution $\mathbf{z} \sim \mathcal{CES}\left(\mathbf{0}, \boldsymbol{\Sigma}, g\right)$, the log-likelihood of a data set $\{\mathbf{z}_k\}$ is given as:

$$\mathcal{L}\left(\{\mathbf{z}_k\}|\boldsymbol{\Sigma}, g\right) = \sum_{k=1}^{K} \log\left(g\left(\mathbf{z}_k^H \boldsymbol{\Sigma}^{-1} \mathbf{z}_k\right)\right) - K \log|\boldsymbol{\Sigma}|. \tag{7}$$

The maximum likelihood estimator of the scatter matrix in this context is the solution of the fixed point equation

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{K} \sum_{k=1}^{K} \psi \left(\mathbf{z}_k^H \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{z}_k \right) \mathbf{z}_k \mathbf{z}_k^H \overset{d}{=} \mathcal{H} \left(\hat{\boldsymbol{\Sigma}} \right), \tag{8}$$

where $\psi(t) = -\phi(t) = -g'(t)/g(t)$. The existence and uniqueness of this solution is subject to conditions on the density generator g and the size of the sample set $\{\mathbf{z}_k\}$ (cf. Theorems 6 and 7 of [15]). When existing, these estimators can be computed through the fixed point iterations $\boldsymbol{\Sigma}_{(n+1)} = \mathcal{H}\left(\boldsymbol{\Sigma}_{(n)}\right)$ that converge to the point $\hat{\boldsymbol{\Sigma}}$ (also Theorem 6 and 7 of [15]). Note that, in practice, the true density generator may not be known or accurately specified. In the robust estimation theory, an M-estimator of the scatter matrix [21, 36] refers to an estimator built using a function $\psi(t)$ that is not necessarily linked to g in (8). Some examples of M-estimators are given in Section VI. These estimators are therefore not maximum likelihoods but are known for their interesting robust and asymptotic properties [15].

An important note is that M-estimators may not be consistent in scale, due to ambiguity discussed in section II-A and exhibited in the simulations of Section VI. A practical way to remove this ambiguity is to focus on the shape matrix estimation by constructing

$$\hat{\mathbf{V}} = \hat{\boldsymbol{\Sigma}} / \sqrt[M]{|\hat{\boldsymbol{\Sigma}}|}, \tag{9}$$

for a given M-estimator (or MLE) of the scatter $\hat{\boldsymbol{\Sigma}}$.

0.2.3 Intrinsic Cramér-Rao bounds

For a good introduction to elementary tools of differential geometry used in this paper, we refer the reader to the appendix of [9], as well as in the footnotes of [1]. For a more detailed coverage of these concepts, one can refer to the standard textbooks [37–41]. The book [42] and manuscript [43] provide algorithmically-oriented introductions to differential geometry.

The intrinsic Cramér-Rao bound extends the traditional Cramér-Rao bound for parameters living in a manifold and for an abitrary chosen Riemannian metric. Indeed, the traditional estimation error (Euclidean distance) is defined through the difference between the true parameter and an estimator, which is not defined intrinsically. To deal with this issue, [1] derived a Cramér-Rao type Theorem for parameters living in a manifold by bounding the expected intrinsic distance between an estimator and the true parameter. Eventually, this Theorem retrieves the well-known inequality " $\mathbf{C} \succcurlyeq \mathbf{F}^{-1}$ ", with $\mathbf{C}$ being the covariance matrix of the estimation error and $\mathbf{F}$ being the Fisher Information Matrix. However, these parameters have a different definition due to the specific nature of the considered objects. The point of this section is to briefly present those definitions and the essential tools needed for the derivation of the contributions. We also refer the reader to the Chapter 6 of [43] and the reference [44], which provide good introductions to the topic.

Theorem 1 *(Fisher information metric, Theorem 1 of [1])*
Let $f(\{\mathbf{z}_k\}|\boldsymbol{\theta})$ be a family of probability density function parameterized by $\boldsymbol{\theta}$ living in a manifold $\mathcal{M}$, $l = \log f$ be the log-likelihood function, and $g_{fim} = [dl \otimes dl]$ ($\otimes$ denotes the tensor product) be the Fisher information metric. Let $\{\boldsymbol{\Omega}\}$ be an element of the tangent space $T_{\boldsymbol{\theta}}\mathcal{M}$ of the manifold $\mathcal{M}$ at point $\boldsymbol{\theta}$. We have the relation

$$g_{fim}\left(\boldsymbol{\Omega}, \boldsymbol{\Omega}\right) = -\mathbb{E}\left[\left. \frac{d^2}{dt^2} l\left(\{\mathbf{z}_k\}|\boldsymbol{\theta} + t\boldsymbol{\Omega}\right)\right|_{t=0} \right]. \tag{10}$$

Let $\{\boldsymbol{\Omega}_i\}$ be a basis $T_{\boldsymbol{\theta}}\mathcal{M}$. The Fisher Information Matrix $\mathbf{F}$ is defined as

$$[\mathbf{F}]_{i,j} = g_{fim}\left(\boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j\right). \tag{11}$$

where $g_{fim}\left(\boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j\right)$ can be obtained from (10) using a polarization formula (cf. (61)).

Definition 1 *(Estimation error)*
Let $\hat{\boldsymbol{\theta}}$ be an estimator of the parameter $\boldsymbol{\theta} \in \mathcal{M}$. The estimation error $\mathbf{X}_{\boldsymbol{\theta}}$ is given by inverse exponential map (or logarithmic map):

$$\mathbf{X}_{\boldsymbol{\theta}} = \exp_{\boldsymbol{\theta}}^{-1} \hat{\boldsymbol{\theta}}. \tag{12}$$

Let the tangent space $T_{\boldsymbol{\theta}}\mathcal{M}$ be endowed with any metric (inner product) $g_{\boldsymbol{\theta}}$ and $\{\boldsymbol{\Omega}_i\}$ be a basis of this space. We have the coordinate vector $\mathbf{x}(\boldsymbol{\theta})$ with entries $[\mathbf{x}\left(\boldsymbol{\theta}\right)]_i = g_{\boldsymbol{\theta}}\left(\mathbf{X}_{\boldsymbol{\theta}}, \boldsymbol{\Omega}_i\right)$, and the squared magnitude of estimation error is

$$\|\mathbf{x}(\boldsymbol{\theta})\|_F^2 = \mathbf{x}(\boldsymbol{\theta})^H\mathbf{x}(\boldsymbol{\theta}) = \left\|\exp_{\boldsymbol{\theta}}^{-1} \hat{\boldsymbol{\theta}}\right\|_{\boldsymbol{\theta}}^2 = d^2(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}), \tag{13}$$

where d is the distance defined w.r.t. the chosen Riemannian metric $g_{\boldsymbol{\theta}}$.

Theorem 2 *(Intrinsic Cramér-Rao Bound, Corollary 2 of [1])*
Let $f(\{\mathbf{z}_k\}|\boldsymbol{\theta})$ be a family of probability density function parameterized by $\boldsymbol{\theta} \in \mathcal{M}$, $l = \log f$ be the log-likelihood function, $g = [dl \otimes dl]$ be the Fisher information metric, $\mathbf{F}$ be the Fisher information Matrix, ∇ be an affine connection on $\mathcal{M}$, and d be the distance associated to $\mathcal{M}$ and chosen Riemannian metric $g_{\boldsymbol{\theta}}$. Assume that $\hat{\boldsymbol{\theta}}$ is an unbiased (cf. definitions 1 and 2 of [1]) estimator of $\boldsymbol{\theta}$, then the covariance of the estimation error $\exp_{\boldsymbol{\theta}}^{-1} \hat{\boldsymbol{\theta}}$ satisfies the matrix inequality

$$\mathbb{E}\left[\left(\exp_{\boldsymbol{\theta}}^{-1} \hat{\boldsymbol{\theta}}\right) \left(\exp_{\boldsymbol{\theta}}^{-1} \hat{\boldsymbol{\theta}}\right)^H\right] \succcurlyeq \mathbf{F}^{-1}, \tag{14}$$

which translates in distance as

$$\mathbb{E}\left[d^2\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\right)\right] \geq \mathrm{Tr}\left\{\mathbf{F}^{-1}\right\}, \tag{15}$$

where d corresponds to the distance associated to the chosen Riemannian metric.

Notice that intrinsic Cramér-Rao bounds in (15) are defined relatively to a Riemannian metric to be chosen, which allows for bounding a distance (performance criterion) that is considered to be meaningful for the addressed estimation problem. From the information geometry [45] perspective, it is generally advocated to consider the distance that is brought by the problem itself, i.e. distance associated to the Fisher information metric from Theorem 1.

0.3 Fisher information metric and natural distance induced by CES distributions

In this section we study the information geometry of the likelihood (7). We derive the Fisher Information Metric and the distance associated to this metric on both $\mathcal{H}_M^{++}$ and $\mathcal{SH}_M^{++}$. These tools will be useful for deriving a natural performance criterion and intrinsic Cramér-Rao lower bounds in the next sections. First, we have the following Theorem:

Theorem 3 *(Fisher Information Metric for CES)*
Let $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ be two vectors of the tangent space of $\mathcal{H}_M^{++}$ at $\mathbf{\Sigma}$, which is $\mathcal{H}_M$. The Fisher Information Metric associated to the likelihood (7) *is:*

$$g_{\mathbf{\Sigma}}^{fim}(\mathbf{\Omega}_1, \mathbf{\Omega}_2) = K\ g_{\mathbf{\Sigma}}^{ces}(\mathbf{\Omega}_1, \mathbf{\Omega}_2), \tag{16}$$

with

$$\begin{aligned} g_{\mathbf{\Sigma}}^{ces}(\mathbf{\Omega}_1, \mathbf{\Omega}_2) &= \alpha \mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}_1\mathbf{\Sigma}^{-1}\mathbf{\Omega}_2\right\} \\ &\quad + \beta \mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}_1\right\}\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}_2\right\}, \end{aligned} \tag{17}$$

and with coefficients α and β defined as

$$\begin{cases} \alpha &= \left(1 - \dfrac{\mathbb{E}\left[\mathcal{Q}_k^2 \phi'(\mathcal{Q}_k)\right]}{M(M+1)}\right) \\ \beta &= \alpha - 1. \end{cases} \tag{18}$$

cf. Appendix A Notice that (17) yields the classical Riemannian metric on $\mathcal{H}_M^{++}$ [46] for $\alpha = 1$ and $\beta = 0$. This also corresponds to the Gaussian case covered in [1] since $\alpha = 1$ and $\beta = 0$ are obtained for the Gaussian density generator $g(t) = \exp(-t)$ (see [29] for the calculation of these coefficients). First, we note that some necessary conditions are to be satisfied by (17) to define a proper Riemannian metric on $\mathcal{H}_M^{++}$:

Proposition 1 *(Positiveness of the Fisher Information Metric) The Fisher Information Metric for CES* (17) *is a Riemannian metric on $\mathcal{H}_M^{++}$ if and only if*

$$\alpha > 0 \quad \text{and} \quad \alpha + M\beta > 0. \tag{19}$$

Cf. Appendix A. Note that from the expressions α and β in (18), this condition is directly reported on the expectation of the second-order modular variate and the density generator g as

$$\mathbb{E}\left[\mathcal{Q}^2\phi^2(\mathcal{Q})\right] > M^2. \tag{20}$$

In the following, we assume that α and β satisfy (19). The natural distance on $\mathcal{H}_M^{++}$ associated to the metric $g_{\mathbf{\Sigma}}^{ces}$ in (17) is obtained by studying the geometry of this Riemannian manifold, which is done in Appendix A and yields:

Theorem 4 *(Distance induced by $g^{ces}_{\mathbf{\Sigma}}$ on $\mathcal{H}_M^{++}$)*
The natural Riemannian distance on $\mathcal{H}_M^{++}$ associated to the metric (17) *is defined, for all $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2 \in \mathcal{H}_M^{++}$, as*

$$d^2_{ces}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \alpha \sum_{i=1}^{M} \log^2 \lambda_i + \beta \left(\sum_{i=1}^{M} \log \lambda_i \right)^2 , \tag{21}$$

where λ_i is the i^{th} eigenvalue of $\mathbf{\Sigma}_1^{-1}\,\mathbf{\Sigma}_2$. It can also be written

$$d^2_{ces}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \alpha \left\| \log(\mathbf{\Sigma}_1^{-1/2}\,\mathbf{\Sigma}_2\,\mathbf{\Sigma}_1^{-1/2}) \right\|_F^2 + \beta \left(\log \left| \mathbf{\Sigma}_1^{-1}\,\mathbf{\Sigma}_2 \right| \right)^2 , \tag{22}$$

Cf. Appendix A.

We now deal with $\mathcal{SH}_M^{++}$, whose tangent space at $\mathbf{\Sigma}$ is

$$T_{\mathbf{\Sigma}}\mathcal{SH}_M^{++} = \{\mathbf{\Omega} \in \mathcal{H}_M : \operatorname{Tr}\{\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}\} = 0\}. \tag{23}$$

It follows that the metric (17) at $\mathbf{\Sigma} \in \mathcal{SH}_M^{++}$ becomes

$$g^{ces}_{\mathbf{\Sigma}}(\mathbf{\Omega}_1, \mathbf{\Omega}_2) = \alpha \operatorname{Tr}\{\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_1\,\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_2\}, \tag{24}$$

for all $\mathbf{\Omega}_1, \mathbf{\Omega}_2 \in T_{\mathbf{\Sigma}}\mathcal{SH}_M^{++}$. As the geodesics on $\mathcal{SH}_M^{++}$ are the same as those on $\mathcal{H}_M^{++}$, the associated distance follows:

Corollary 1 *(Distance induced by $g^{ces}_{\mathbf{\Sigma}}$ on $\mathcal{SH}_M^{++}$)*
The natural Riemannian distance on $\mathcal{SH}_M^{++}$ associated to the metric (17) *is defined, for all $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2 \in \mathcal{SH}_M^{++}$, as*

$$d^2_{sp-ces}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \alpha \left\| \log(\mathbf{\Sigma}_1^{-1/2}\,\mathbf{\Sigma}_2\,\mathbf{\Sigma}_1^{-1/2}) \right\|_F^2 . \tag{25}$$

Note that d_{ces} corresponds to a scaled natural distance plus an additional term that comes from the integration of factors in β in (17) along the geodesic. However, these terms cannot define alternate Riemannian metric and distance by themselves, as $\alpha = 0$ does not satisfy Proposition 1. On the other hand, d_{sp-ces} corresponds to a scaled natural distance for any β.

0.4 Intrinsic Cramér-Rao Bounds on scatter

In this Section we derive intrinsic Cramér-Rao bounds for the problem of scatter matrix estimation in CES distributions under different metrics, hence performance bounds on different distances. Once a distance is chosen to evaluate the performance in terms of estimation accuracy, the Cramér-Rao bound is obtained by performing the following steps:

a) Selecting a basis $\{\mathbf{\Omega}_i\}$ of $\mathcal{H}_M$ (the tangent space of $\mathcal{H}_M^{++}$ at $\mathbf{\Sigma}$) that is orthonormal w.r.t. the metric on $\mathcal{H}_M^{++}$ that is associated to this distance.

$b)$ Computing the elements of the Fisher Information Matrix with this basis, according to Theorem 1.

$c)$ Inverting the FIM and applying Theorem 2.

These operations correspond to the steps described in the frame "Computation of the intrinsic FIM and CRB" in [1].

0.4.1 Euclidean Metric

We first recall that the Euclidean metric for $\mathcal{H}_M^{++}$ and associated distance are:

$$g^{Eucl}\left(\mathbf{\Omega}, \mathbf{\Omega}\right) = \mathrm{Tr}\left\{\mathbf{\Omega}^2\right\}, \tag{26}$$

$$d^2_{Eucl}\left(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2\right) = \left\|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\right\|_F^2. \tag{27}$$

Consider the following basis of the tangent space of $\mathcal{H}_M^{++}$ at $\mathbf{\Sigma}$, orthonormal w.r.t. to the inner product g^{Eucl}:

1. $\mathbf{\Omega}_{ii}^{Eucl}$ is an n by n symmetric matrix whose ith diagonal element is one, zeros elsewhere
2. $\mathbf{\Omega}_{ij}^{Eucl}$ is an n by n symmetric matrix whose ijth and jith elements are both $2^{-1/2}$, zeros elsewhere.
3. $\mathbf{\Omega}_{ij}^{h-Eucl}$ is an n by n Hermitian matrix whose ijth element is $2^{-1/2}\sqrt{-1}$, and jith element is $-2^{-1/2}\sqrt{-1}$, zeros elsewhere ($i < j$).

To shorten notations, we simply denote this basis $\{\mathbf{\Omega}_i^{Eucl}\}$, for $i \in [\![1, M^2]\!]$, where the M^2 elements are corresponding to the $\mathbf{\Omega}_{ij}^{\{h\}}$ that are ordered following items 1) 2) 3). The distance between an estimator and its true value is obtained as the summed squared errors on the coordinates in this basis.

Theorem 5 *(Euclidean Cramér-Rao bound on the scatter)*
Let $\hat{\mathbf{\Sigma}}$ be an estimator of the scatter matrix built from a data set $\{\mathbf{z}_k\}$ i.i.d. distributed according $\mathbf{z} \sim \mathcal{CES}\left(\mathbf{0}, \mathbf{\Sigma}, g\right)$. The Cramér-Rao bound on the Euclidean distance between $\hat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$ is

$$\mathbb{E}\left[d^2_{Eucl}\left(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}\right)\right] \geq \mathrm{Tr}\left\{\mathbf{F}_{Eucl}^{-1}\right\}, \tag{28}$$

with

$$\begin{aligned} \left[\mathbf{F}_{Eucl}\right]_{i,j} &= K\alpha\, \mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}_i^{Eucl}\mathbf{\Sigma}^{-1}\mathbf{\Omega}_j^{Eucl}\right\} \\ &\quad + K\beta\, \mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}_i^{Eucl}\right\} \mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}_j^{Eucl}\right\}. \end{aligned} \tag{29}$$

Let $\{\mathbf{\Omega}_i^{Eucl}\}_{i \in 1 \ldots M^2}$ be the canonical basis, and $g_{\mathbf{\Sigma}}^{fim}$ be the Fisher information metric defined in Theorem 3. The result is a direct application of Theorem 2.

Remark that this corresponds well to the Euclidean Cramér-Rao bounds obtained for several distributions in [26–29]. Also notice that we retrieve the same result as Theorem 5 of [1] for the special case of Gaussian distribution, i.e. $\alpha = 1$ and $\beta = 0$.

0.4.2 Natural Metric

Recall that the natural metric and associated distance are

$$g_{\mathbf{\Sigma}}^{nat}\left(\mathbf{\Omega},\mathbf{\Omega}\right)=\mathrm{Tr}\left\{\left(\mathbf{\Sigma}^{-1}\mathbf{\Omega}\right)^{2}\right\}, \tag{30}$$

$$d_{nat}^{2}\left(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2\right)=\left\|\log(\mathbf{\Sigma}_1^{-1/2}\,\mathbf{\Sigma}_2\,\mathbf{\Sigma}_1^{-1/2})\right\|_F^2. \tag{31}$$

An orthonormal basis for the tangent space of $\mathcal{H}_M^{++}$ at $\mathbf{\Sigma}$ w.r.t. $g_{\mathbf{\Sigma}}^{nat}$ can be obtained by coloring the canonical basis of previous section as:

$$\mathbf{\Omega}_i^{nat}=\mathbf{\Sigma}^{1/2}\mathbf{\Omega}_i^{Eucl}\mathbf{\Sigma}^{1/2}. \tag{32}$$

We denote this basis $\{\mathbf{\Omega}_i^{nat}\}_{i\in 1\dots M^2}$. The distance between an estimator and the true scatter matrix is obtained as the summed squared errors on the coordinates in this basis (see (84) to (87) in [1]). This distance is subject to the following bound:

Theorem 6 *(Natural Cramér-Rao bound on the scatter)*
Let $\hat{\mathbf{\Sigma}}$ be an estimator of the scatter matrix built from a data set $\{\mathbf{z}_k\}$ i.i.d. distributed according $\mathbf{z}\sim\mathcal{CES}\left(\mathbf{0},\mathbf{\Sigma},g\right)$. The Cramér-Rao bound on the natural distance between $\hat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$ is

$$\mathbb{E}\left[d_{nat}^{2}\left(\hat{\mathbf{\Sigma}},\mathbf{\Sigma}\right)\right]\geq\frac{M^2-1}{K\alpha}+(K(\alpha+M\beta))^{-1}. \tag{33}$$

Let $\{\mathbf{\Omega}_i^{nat}\}_{i\in 1\dots M^2}$ be the tangent space basis defined in (32), and $g_{\mathbf{\Sigma}}^{fim}$ be the Fisher information metric defined in Theorem 3. Notice that $\mathbf{\Omega}_i^{nat}=\mathbf{\Sigma}^{1/2}\mathbf{\Omega}_i^{Eucl}\mathbf{\Sigma}^{1/2}$, which simplifies the entries of the Fisher information matrix as:

$$\begin{aligned} g_{\mathbf{\Sigma}}^{fim}\left(\mathbf{\Omega}_i,\mathbf{\Omega}_j\right) &= K\alpha\mathrm{Tr}\left\{\mathbf{\Omega}_i^{Eucl}\mathbf{\Omega}_j^{Eucl}\right\} \\ &\quad + K\beta\mathrm{Tr}\left\{\mathbf{\Omega}_i^{Eucl}\right\}\mathrm{Tr}\left\{\mathbf{\Omega}_j^{Eucl}\right\}. \end{aligned} \tag{34}$$

Hence, from the relations

$$\mathrm{Tr}\left\{\mathbf{\Omega}_i^{Eucl}\mathbf{\Omega}_j^{Eucl}\right\}=\delta_{i,j}, \tag{35}$$

and

$$\mathrm{Tr}\left\{\mathbf{\Omega}_i^{Eucl}\right\}\mathrm{Tr}\left\{\mathbf{\Omega}_j^{Eucl}\right\}=\begin{cases}1 \text{ if } (i,j)\in[\![1,n]\!]^2\\ \quad 0 \text{ otherwise,}\end{cases} \tag{36}$$

we obtain the Fisher information matrix as

$$\mathbf{F}_{nat}=K\alpha\mathbf{I}_{M^2}+K\beta\begin{bmatrix}\mathbf{1}_{M\times M} & \mathbf{0}_{1\times M(M-1)}\\ \mathbf{0}_{M(M-1)\times 1} & \mathbf{0}_{M(M-1)\times M(M-1)}\end{bmatrix},$$

which is expressed as $\mathbf{F}_{nat}=K\alpha\mathbf{I}+KM\beta\mathbf{v}_{fim}\mathbf{v}_{fim}^H$ with unitary vector $\mathbf{v}_{fim}=1/\sqrt{M}\left[\,\mathbf{1}_M\mid\mathbf{0}_{M(M-1)}\,\right]$, i.e. $\mathbf{v}_{fim}^H\mathbf{v}_{fim}=1$. Hence $\mathbf{F}_{nat}^{-1}$ can be obtained by the Sherman-Morrison formula, or its vector of eigenvalues can be directly identified as $K^{-1}\left[(\alpha+M\beta)^{-1},\alpha^{-1},\dots,\alpha^{-1}\right]$ and summed to obtain its trace. Theorem 2 is applied to conclude.

0.4.3 CES-Fisher Information Metric

Recall that the CES-Fisher information metric and associated distance are given in (17) and (21) respectively. We denote $\{\boldsymbol{\Omega}_i^{ces}\}_{i\in 1\ldots M^2}$, a basis that is orthonormal w.r.t. to the metric (inner product) $g_{\boldsymbol{\Sigma}}^{ces}$. Closed-form expressions of this basis are not needed for the developments, but it can be constructed using Gram-Schmidt orthogonalization process. The distance between an estimator and the true scatter matrix is obtained as the summed squared errors on the coordinates in this basis. This distance leads to the following bound:

Theorem 7 *(CES-Fisher Cramér-Rao bound on the scatter)*
Let $\hat{\boldsymbol{\Sigma}}$ be an estimator of the scatter matrix built from a data set $\{\mathbf{z}_k\}$ i.i.d. distributed according $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g)$. The Cramér-Rao bound on the CES-Fisher distance between $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}$ is

$$\mathbb{E}\left[d_{ces}^2\left(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}\right)\right] \geq M^2/K. \tag{37}$$

Let $\{\boldsymbol{\Omega}_i^{ces}\}_{i\in 1\ldots M^2}$ be the tangent space basis defined, orthonormal w.r.t. to the metric (inner product) $g_{\boldsymbol{\Sigma}}^{ces}$. Let $g_{\boldsymbol{\Sigma}}^{fim}$ be the Fisher information metric defined in Theorem 3. Notice that $g_{\boldsymbol{\Sigma}}^{fim} = K g_{\boldsymbol{\Sigma}}^{ces}$, so the Fisher Information Matrix is, by construction (orthonormality) equal to $\mathbf{F}_{ces} = K\mathbf{I}_{M^2}$. The trace of its inverse is therefore M^2/K and the proof is concluded by applying Theorem 2. Notice that in the Gaussian case, Theorems 7 and 6 become identical since $\alpha = 1$ and $\beta = 0$.

0.5 Intrinsic Cramér-Rao Bounds on shape

In this section we derive Intrinsic Cramér-Rao bounds for the problem of shape matrix estimation in CES distributions under different metrics, hence performance bounds on different distances. Overall, we follow the same steps $a)$, $b)$ and $c)$ as described in the beginning of Section IV. However, the shape matrix parameter lives in the Manifold $\mathcal{SH}_M^{++}$, so the considered tangent space has to be modified accordingly. At the point $\boldsymbol{\Sigma}$, the tangent space of $\mathcal{SH}_M^{++}$ is defined in (23).

0.5.1 Euclidean Metric

First recall that the Euclidean metric and distances are given respectively in (26) and (27). A practical alternate formulation is that $T_{\boldsymbol{\Sigma}}\mathcal{SH}_M^{++}$ is the complementary of the normal space at $\boldsymbol{\Sigma}$, denoted $N_{\boldsymbol{\Sigma}}\mathcal{SH}_M^{++}$. When using g^{Eucl} in (26) as inner product, this space is defined as

$$N_{\boldsymbol{\Sigma}}\mathcal{SH}_M^{++} = \left\{\lambda\boldsymbol{\Sigma}^{-1}, \lambda \in \mathbb{R}\right\}, \tag{38}$$

so $T_{\boldsymbol{\Sigma}}\mathcal{SH}_M^{++}$ corresponds to the space of symmetric matrices deprived from the line $\lambda\boldsymbol{\Sigma}^{-1}$, $\lambda \in \mathbb{R}$. Therefore, the main trick to obtain an orthonormal basis of $T_{\boldsymbol{\Sigma}}\mathcal{SH}_M^{++}$ will be to construct one of $\mathcal{H}_M$ where $\lambda\boldsymbol{\Sigma}$ appears as the first element. Extracting the $M^2 - 1$ other elements of this set will lead to the desired basis.

Consider the basis $\left\{\boldsymbol{\Omega}_i^{Eucl}\right\}$ defined in Section IV-A, augmented with the element $\boldsymbol{\Sigma}^{-1}$ as $\left\{\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Omega}_1^{Eucl}, \ldots, \boldsymbol{\Omega}_{M^2}^{Eucl}\right\}$. Then, applying a Gram-Schmidt orthonormalization process with the scalar product defined in (26) on this set leads to $\left\{\lambda\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Omega}_1^{sp-Eucl}, \ldots, \boldsymbol{\Omega}_{M^2-1}^{sp-Eucl}, \mathbf{0}\right\}$ (with appropriate normalization λ). Now, extracting M^2-1 elements (excluding $\lambda\boldsymbol{\Sigma}^{-1}$ and $\mathbf{0}$) leads to an orthonormal basis of $T_{\boldsymbol{\Sigma}}\mathcal{SH}_M^{++}$ w.r.t. g^{Eucl} denoted

$$\left\{\boldsymbol{\Omega}_1^{sp-Eucl}, \ldots, \boldsymbol{\Omega}_{M^2-1}^{sp-Eucl}\right\}. \tag{39}$$

Theorem 8 *(Euclidean Cramér-Rao bound on the shape)*
Let $\hat{\mathbf{V}}$ be an estimator of the shape matrix $\mathbf{V} \in \mathcal{SH}_M^{++}$ built from a data set $\{\mathbf{z}_k\}$ i.i.d. distributed according $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g)$ with $\boldsymbol{\Sigma} = \sigma^2\mathbf{V}$ and equivalent distribution $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{V}, \tilde{g})$ (cf. section II-A). The Cramér-Rao bound on the Euclidean distance between $\hat{\mathbf{V}}$ and $\mathbf{V}$ is

$$\mathbb{E}\left[d_{Eucl}^2\left(\hat{\mathbf{V}}, \mathbf{V}\right)\right] \geq \operatorname{Tr}\left\{\mathbf{F}_{sp-Eucl}^{-1}\right\}, \tag{40}$$

with

$$\left[\mathbf{F}_{sp-Eucl}\right]_{i,j} = K\alpha \operatorname{Tr}\left\{\mathbf{V}^{-1}\boldsymbol{\Omega}_i^{sp-Eucl}\mathbf{V}^{-1}\boldsymbol{\Omega}_j^{sp-Eucl}\right\}. \tag{41}$$

for $i, j \in [\![1, M^2-1]\!]$ and with α from (18) using $\tilde{g}$.

Let $\{\boldsymbol{\Omega}_i^{sp-Eucl}\}_{i\in 1\ldots M^2-1}$ be the basis in (39), and $g_{\boldsymbol{\Sigma}}^{fim}$ be the Fisher information metric defined in Theorem 3. The result is a direct application of Theorem 2. Note that the terms in β are all zero due to the definition of the tangent space $T_{\boldsymbol{\Sigma}}\mathcal{SH}_M^{++}$ in (23).

Note that this theorem allows to compute the Cramér-Rao lower bound on the shape matrix parameter in a practical way and without requiring a subtle parameterization that ensures unit determinant. This is, to the best of our knowledge, a new result even for the Euclidean setting.

0.5.2 Natural and CES-Fisher Information Metric

First, recall that the natural metric and distances are given respectively in (30) and (31). Also note that from Corollary 1, the CES-Fisher distance corresponds to a scaled natural distance on the space $\mathcal{SH}_M^{++}$. Hence the following analysis holds for both Natural and CES-Fisher distances up to a scale factor in definition of the estimation error. When using $g_{\boldsymbol{\Sigma}}^{nat}$ in (30) as inner product, the normal space at $\boldsymbol{\Sigma}$ is defined as

$$N_{\boldsymbol{\Sigma}}\mathcal{SH}_M^{++} = \{\lambda\boldsymbol{\Sigma}, \lambda \in \mathbb{R}\}. \tag{42}$$

To obtain an orthonormal basis of the tangent space, we follow the steps of Section V-A, except that $\boldsymbol{\Sigma}$ is used to augment the initial basis, and $g_{\boldsymbol{\Sigma}}^{nat}$ in (30) is used as inner product to perform the Gram-Schmidt orthonormalization process. This leads to an orthonormal basis of $T_{\boldsymbol{\Sigma}}\mathcal{SH}_M^{++}$ w.r.t. $g_{\boldsymbol{\Sigma}}^{nat}$ denoted

$$\left\{\boldsymbol{\Omega}_1^{sp-nat}, \ldots, \boldsymbol{\Omega}_{M^2-1}^{sp-nat}\right\}. \tag{43}$$

Theorem 9 *(Natural Cramér-Rao bound on the shape)*
Let $\hat{\mathbf{V}}$ be an estimator of the shape matrix $\mathbf{V} \in \mathcal{SH}_M^{++}$ built from a data set $\{\mathbf{z}_k\}$ i.i.d. distributed according $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g)$ with $\mathbf{\Sigma} = \sigma^2 \mathbf{V}$ and equivalent distribution $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{V}, \tilde{g})$ (cf. section II-A). The Cramér-Rao bound on the natural Riemannian distance between $\hat{\mathbf{V}}$ and $\mathbf{V}$ is

$$\mathbb{E}\left[d_{nat}^2\left(\hat{\mathbf{V}}, \mathbf{V}\right)\right] \geq \frac{M^2 - 1}{K\alpha} \tag{44}$$

with α from (18) *using $\tilde{g}$.*

Let $\{\mathbf{\Omega}_i^{sp-nat}\}$ be the basis in (43), and $g_{\mathbf{\Sigma}}^{fim}$ be the Fisher information metric defined in Theorem 3. The entries of the Fisher Information Matrix are

$$g_{\mathbf{\Sigma}}^{fim}\left(\mathbf{\Omega}_i^{sp-nat}, \mathbf{\Omega}_j^{sp-nat}\right) = K\alpha\, \delta_{i,j} \tag{45}$$

due to the orthonormality of $\{\mathbf{\Omega}_i^{sp-nat}\}$ w.r.t. g^{nat} (the terms in α), and thanks to the fact that terms in β are all zero from the definition of the tangent space $T_{\mathbf{\Sigma}}\mathcal{SH}_M^{++}$ in (23). The Fisher information matrix is therefore $\mathbf{F}_{sp-nat} = K\alpha \mathbf{I}_{M^2-1}$ whose the trace of inverse reads directly. The proof is concluded by applying Theorem 2.

0.6 Simulations

0.6.1 Student t-distribution

In this section, the theoretical results of previous sections will be illustrated for the multivariate Student t-distribution. The multivariate Student distribution with $d \in \mathbb{N}^*$ degree of freedom is obtained for the CES representation $\mathbf{z} \sim (\mathbf{0}, \mathbf{\Sigma}, g_d)$ with

$$g_d(t) = \left(1 + d^{-1}t\right)^{-(d+M)}, \tag{46}$$

and the second-order modular variate is distributed as $\mathcal{Q} \overset{d}{=} \mathbb{C}_{\mathcal{X}_M^2} / \mathbb{C}_{\mathcal{X}_d^2/d}$ where $\mathbb{C}_{\mathcal{X}_x^2}$ denotes the Chi-squared distribution with x degrees of freedom. Hence $\mathcal{Q}$ follows a scaled $\mathcal{F}$-distribution. We have

$$\phi(t) = -\frac{d+M}{d+t}, \tag{47}$$

and the expectation

$$\mathbb{E}\left[\mathcal{Q}^2 \phi^2(\mathcal{Q})\right] = \frac{(d+M)M(M+1)}{d+M+1}, \tag{48}$$

leading to the coefficients of the metric $g_{\mathbf{\Sigma}}^{ces}$ in (17) as

$$\begin{cases} \alpha &= \dfrac{d+M}{d+M+1} \\ \beta &= \dfrac{-1}{d+M+1}. \end{cases} \tag{49}$$

Notice that the condition of Proposition 1, $\alpha + M\beta > 0$, is valid $\forall d \in \mathbb{N}^*$, so the Fisher Information Metric is always properly defined for this distribution.

0.6.2 Studied M-estimators

We consider the following estimators of the scatter:

- SCM: the usual Sample Covariance Matrix, defined as $\hat{\mathbf{\Sigma}}_{SCM} = K^{-1} \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H$.
- MLE: The estimator defined in (8) using the appropriate function $\psi(t) = -\phi(t)$, with ϕ in (47).
- Mismatched MLE: the M-estimator identically to the MLE, except that the used parameter d different from the true parameter. Here, $d = 10$ is set regardless of the underlying distribution.
- Tyler's M-estimator, defined as in (8) with $\psi(t) = M/Kt$. Note that this estimator is unique up to a scaling factor so it will be considered only for shape estimation.

For all this estimates, the corresponding estimators of the shape are build using the normalization in (9).

0.6.3 Simulations results

In this section, the scatter matrix is built as a Toeplitz matrix $[\mathbf{\Sigma}_T]_{i,j} = \rho^{|i-j|}$ with $\rho = 0.9\sqrt{1/2}\,(1+i)$. This matrix is then normalized so that the scatter and shape matrices are equal in this setting. For samples distributed as $\mathbf{z} \sim \mathcal{CES}\left(\mathbf{0}, \mathbf{\Sigma}_T, g_d\right)$, we study the performance of the different estimators of scatter and shape versus the number of samples K (evaluated on 10^4 Monte-Carlo simulations). These performances are evaluated through the mean squared distances corresponding to g^{Eucl}, $g^{nat}_{\mathbf{\Sigma}}$ and $g^{ces}_{\mathbf{\Sigma}}$ and are compared to the corresponding Cramér-Rao lower bounds from Section IV.

Figure 1 displays the results on scatter matrix estimation for a Student t-distribution with $d = 100$ degrees of freedom. Notice that in this case the data almost follow a Gaussian distribution (it is usually admitted that $d > 30$ allows to assume Gaussianity of the data). In this setting $\hat{\mathbf{\Sigma}}_{MLE} \simeq \hat{\mathbf{\Sigma}}_{SCM}$ so these estimators reach similar performances. For all performance measurements (different distance), the mismatched MLE appears not efficient at high sample support, which is due to the bias induced on the scale. Also, $\alpha \simeq 1$ and $\beta \simeq 0$, so $g^{nat}_{\mathbf{\Sigma}}$ and $g^{ces}_{\mathbf{\Sigma}}$ generate almost identical distances and corresponding bounds, as observed in Figure 1. Interestingly, as noted in [1], these performance criteria show that the studied estimators are not efficient at low sample support. The natural metric is able to reflect some empirical results in terms of application (the Sample Covariance Matrix is known to provide an inaccurate estimation at low sample support), while the Euclidean metric is apparently not, i.e. the Cramér-Rao bound and MSE on the Euclidean metric appear non-informative here.

Figure 2 displays the same results for a Student t-distribution with $d = 2$ degrees of freedom. Here, the distribution is heavy tailed and the SCM, as well as the mismatched MLE, fail to provide an accurate estimator of the scatter matrix, as observed in Figure 2. In this case, the study of the Euclidean metric reveals that the Maximum Likelihood Estimator is not efficient at low sample support, however it converges to the bound as K grows. We notice that the convergence towards this regime appears to be slower

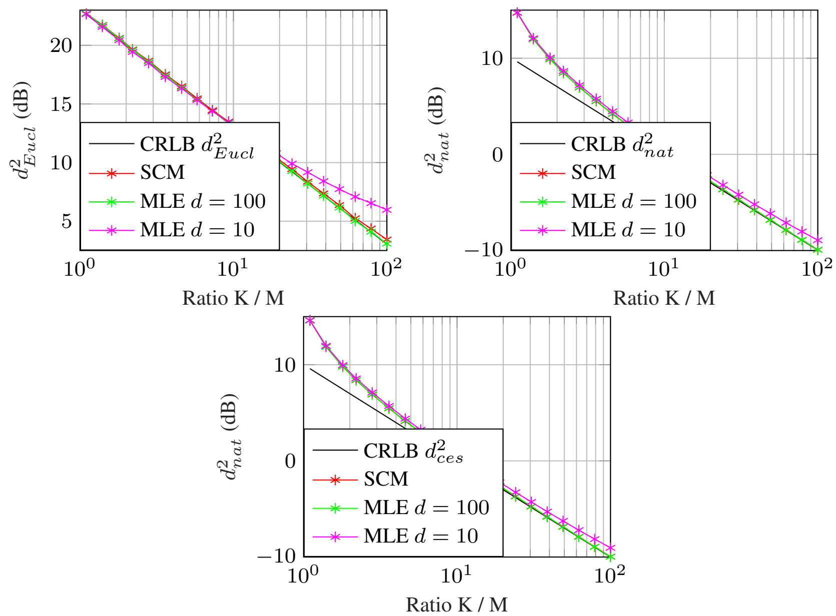

Figure 1: (from top to bottom) Euclidean, Natural, CES-Fisher CRLB and mean squared distance scatter matrix for t-distribution versus K/M. $M = 10$, $d = 100$ (close to Gaussian case).

through the study of the natural and CES-Fisher metric, which may be an interesting point in order to quantify the number of samples needed to achieve good performance in terms of application purpose.

Figures 3 and 4 display the results for the problem of shape matrix estimation for the same configurations as in Figures 1 and 2. The interest of this setting and the derived intrinsic Cramér-Rao bounds is that they allow to draw a meaningful comparison of different M-estimators using both Euclidean and Natural distance, regardless of the scaling ambiguities inherent to CES distributions. Such comparison is indeed relevant when the process of interest is not sensitive to scale (e.g. for adaptive filtering). Here, both distance reveal that all the studied shape matrix estimators are not efficient at low sample support. We also notice that M-estimators such as the mismatched MLE and Tyler's estimator appear here close to the MLE in terms of performance for the problem of shape estimation. However, this is not the case for the SCM if the distribution is not Gaussian, as seen in Figure 4.

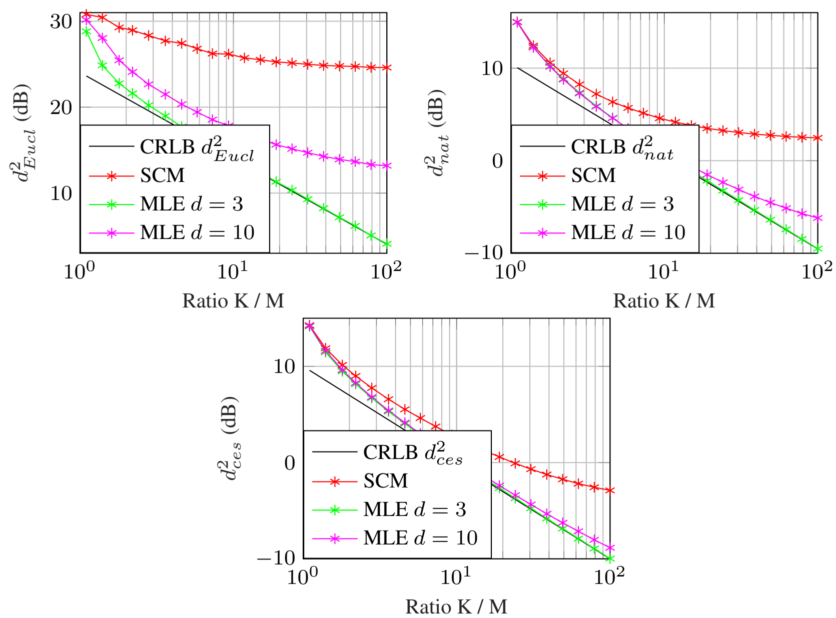

Figure 2: (from top to bottom) Euclidean, Natural, CES-Fisher CRLB and mean squared distance on scatter matrix for t-distribution versus K/M. $M = 10$, $d = 3$.

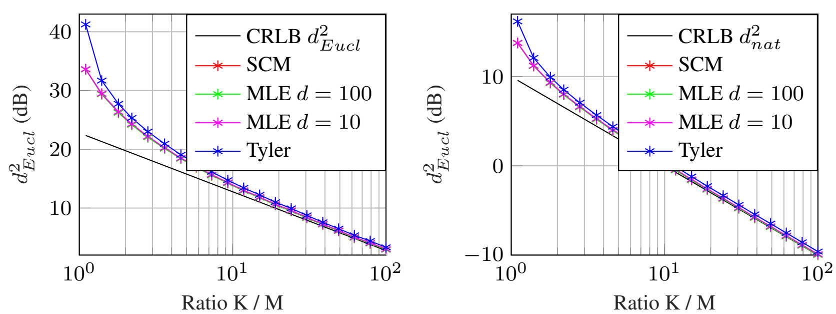

Figure 3: (from top to bottom) Euclidean and Natural CRLB and Mean Squared Distance on shape versus K/M. $M = 10$, $d = 100$ (close to Gaussian case).

0.7 Conclusion

This paper derived intrinsic Cramér-Rao bounds for the problem of scatter and shape matrices estimation from samples following a CES distribution. The intrinsic approach allowed to obtain performance bounds on three different distances (Euclidean, natural Riemannian and CES-Fisher). An interesting point is that the study of these three re-

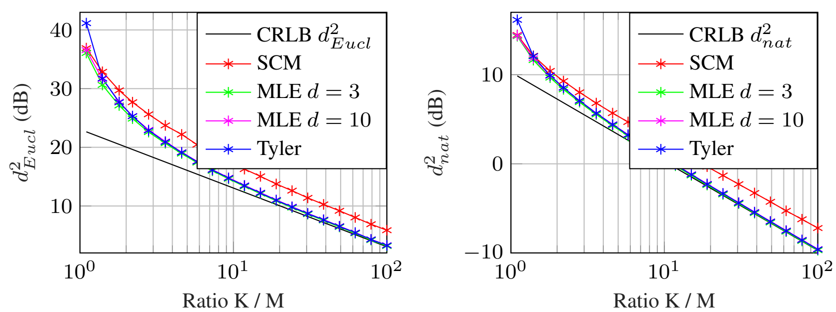

Figure 4: (from top to bottom) Euclidean and Natural CRLB and Mean Squared Distance on shape versus K/M. $M = 10$, $d = 3$.

sults can reveal hidden properties of estimators. Therefore, the obtained Cramér-Rao bounds are useful tools for a complete analysis. In this scope, the proposed results on shape matrices allow to draw performance bounds for the comparison of all M-estimators, regardless of scaling ambiguities inherent to CES distributions. On a side note, the Fisher-CES metric presented in this paper allows to build generalized Riemannian distances while preserving the natural geodesic of $\mathcal{S}_M^{++}$. This tool seems therefore interesting for building regularized estimators in the vein of [22–25].

0.8 Proofs of Section III

Proof of Theorem 3

The Fisher Information Metric is obtained according to Theorem 1 as

$$g_{\mathbf{\Sigma}}^{fim}\left(\mathbf{\Omega}, \mathbf{\Omega}\right) = -\mathbb{E}\left[\left.\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{L}\left(\{\mathbf{z}_k\}|\mathbf{\Sigma} + t\partial\mathbf{\Omega}, g\right)\right|_{t=0}\right]. \tag{50}$$

First, recall that the log-likelihood of the sample set is

$$\mathcal{L}\left(\{\mathbf{z}_k\}|\mathbf{\Sigma}, g\right) = \sum_{k=1}^{K}\log\left(g\left(\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right)\right) - K\log|\mathbf{\Sigma}|, \tag{51}$$

where $\mathbf{Z}_k = \mathbf{z}_k\mathbf{z}_k^H$. We have the following Taylor expansions of order two around $\mathbf{\Sigma}$:

$$\log|\mathbf{\Sigma} + t\mathbf{\Omega}| = \log|\mathbf{\Sigma}| + \mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}t\mathbf{\Omega}\right\} - \frac{1}{2}\mathrm{Tr}\left\{\left(\mathbf{\Sigma}^{-1}t\mathbf{\Omega}\right)^2\right\} + \dots \tag{52}$$

and

$$\begin{aligned}\log\left(g\left(\mathrm{Tr}\left\{\left(\mathbf{\Sigma} + t\mathbf{\Omega}\right)^{-1}\mathbf{Z}_k\right\}\right)\right) = \;& \log\left(g\left(\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right)\right)\\ & - \mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}t\mathbf{\Omega}\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\psi\left(\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right)\\ & + \mathrm{Tr}\left\{\left(t\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\mathbf{Z}_k\mathbf{\Sigma}^{-1}\right\}\psi\left(\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right)\\ & + \frac{1}{2}\mathrm{Tr}^2\left\{\mathbf{\Sigma}^{-1}t\mathbf{\Omega}\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\psi'\left(Tr\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right) + \dots\end{aligned} \tag{53}$$

By removing the higher order terms we obtain

$$
\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{L}\left(\{\mathbf{z}_k\}|\mathbf{\Sigma}+t\mathbf{\Omega}\right)\Big|_{t=0} = \; & K\mathrm{Tr}\left\{\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\right\} \\
& + 2\sum_{k=1}^{K}\mathrm{Tr}\left\{\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\mathbf{Z}_k\mathbf{\Sigma}^{-1}\right\}\phi\left(Tr\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right) \\
& + \sum_{k=1}^{K}\mathrm{Tr}^2\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\phi'\left(Tr\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right).
\end{aligned} \tag{54}
$$

In order to compute the expectations, we recall that $\mathbf{Z}_k = \mathbf{z}_k\mathbf{z}_k^H$ and that $\mathbf{z}_k$ has the stochastic representation $\mathbf{z}_k \overset{d}{=} \sqrt{\mathcal{Q}_k}\ \mathbf{\Sigma}^{1/2}\ \mathbf{u}_k$. This allows us some simplifications since $\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\} = \mathcal{Q}_k$, $\mathbf{u}_k^H\mathbf{u}_k = 1$, and since that $\mathbf{u}_k$ and $\mathcal{Q}_k$ are independent (allowing to split the expectations). Hence we have for the first term:

$$
\begin{aligned}
& \mathbb{E}\left[\mathrm{Tr}\left\{\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\mathbf{Z}_k\mathbf{\Sigma}^{-1}\right\}\phi\left(\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right)\right] \\
& = \mathbb{E}\left[\mathrm{Tr}\left\{\mathbf{\Sigma}^{H/2}\mathbf{\Sigma}^{-1}\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\mathbf{\Sigma}^{1/2}\mathbf{u}_k\mathbf{u}_k^H\right\}\right]\mathbb{E}\left[\mathcal{Q}_k\phi\left(\mathcal{Q}_k\right)\right] \\
& = -\mathrm{Tr}\left\{\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\right\},
\end{aligned} \tag{55}
$$

where we used $\mathbb{E}\left[\mathbf{u}_k\mathbf{u}_k^H\right] = \mathbf{I}_M/M$ (since $\mathbf{u}_k \sim \mathcal{U}(\mathbb{C}\mathcal{S}^M)$), and (4) to obtain the result

$$
\mathbb{E}\left[\mathcal{Q}_k\phi\left(\mathcal{Q}_k\right)\right] = -M. \tag{56}
$$

The second expectation is obtained by the same method as

$$
\begin{aligned}
& \mathbb{E}\left[\mathrm{Tr}^2\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\phi'\left(\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right)\right] \\
& = \mathbb{E}\left[\left(\mathbf{u}_k^H\mathbf{\Sigma}^{-H/2}\mathbf{\Omega}\mathbf{\Sigma}^{-1/2}\mathbf{u}_k\right)^2\right]\mathbb{E}\left[\mathcal{Q}_k^2\phi'\left(\mathcal{Q}_k\right)\right] \\
& = \frac{\mathbb{E}\left[\mathcal{Q}_k^2\phi'\left(\mathcal{Q}_k\right)\right]}{M(M+1)}\left(\mathrm{Tr}^2\left\{\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right\} + \mathrm{Tr}\left\{\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\right\}\right),
\end{aligned} \tag{57}
$$

where we used the relation from [29], giving

$$
\mathbb{E}\left[\left(\mathbf{u}_k^H\mathbf{B}\mathbf{u}_k\right)^2\right] = \frac{\mathrm{Tr}\left\{\mathbf{B}^2\right\} + \mathrm{Tr}^2\left\{\mathbf{B}\right\}}{M(M+1)}, \tag{58}
$$

for an arbitrary constant matrix $\mathbf{B}$ and $\mathbf{u}_k \sim \mathcal{U}\left(\mathbb{C}\mathcal{S}^M\right)$. Eventually, by plugging (55) and (57) into (50) and (54), the Fisher Information Metric is given as:

$$
g_{\mathbf{\Sigma}}^{fim}\left(\mathbf{\Omega},\mathbf{\Omega}\right) = K\alpha\mathrm{Tr}\left\{\left(\mathbf{\Sigma}^{-1}\mathbf{\Omega}\right)^2\right\} + K\beta\mathrm{Tr}^2\left\{\mathbf{\Omega}\mathbf{R}^{-1}\right\}, \tag{59}
$$

with coefficients α and β defined in (18). Notice that $\beta = \alpha - 1$. Also, some manipulations with $\phi'(t) = g''(t)/g(t) - \phi^2(t)$, (4) and (56) allow to show that

$$
M(M+1) - \mathbb{E}\left[\mathcal{Q}_k^2\phi'\left(\mathcal{Q}_k\right)\right] = \mathbb{E}\left[\mathcal{Q}_k^2\phi^2\left(\mathcal{Q}_k\right)\right], \tag{60}
$$

which is consistent with the coefficients obtained in the parametric case [29]. To obtain the metric we now use the polarization formula

$$
\begin{aligned}
g_{\mathbf{\Sigma}}^{fim}\left(\mathbf{\Omega}_1,\mathbf{\Omega}_2\right) \; = \; & \frac{1}{4}\left[g_{\mathbf{\Sigma}}^{fim}\left(\mathbf{\Omega}_1+\mathbf{\Omega}_2,\mathbf{\Omega}_1+\mathbf{\Omega}_2\right)\right. \\
& \left. - g_{\mathbf{\Sigma}}^{fim}\left(\mathbf{\Omega}_1-\mathbf{\Omega}_2,\mathbf{\Omega}_1-\mathbf{\Omega}_2\right)\right],
\end{aligned} \tag{61}
$$

which, after some expansions and simplifications leads to the conclusion of the proof.

Proof of Proposition 1

It is readily checked that for all $\mathbf{\Sigma} \in \mathcal{S}_M^{++}$ the function (17) is symmetric and bilinear. It remains to determine whether it is positive-definite. Let $\mathbf{\Sigma} \in \mathcal{S}_M^{++}$ and $\mathbf{\Omega} \in \mathcal{S}_M$, and let $\mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^T$ be the eigenvalue decomposition of $\mathbf{\Sigma}^{-1/2}\,\mathbf{\Omega}\,\mathbf{\Sigma}^{-1/2}$. One can first check that we need $\alpha > 0$ because the term on the right can be canceled for $\mathbf{\Omega}$ different from $\mathbf{0}$. We have

$$\begin{aligned} g_{\mathbf{\Sigma}}^{ces}(\mathbf{\Omega},\mathbf{\Omega}) &= \alpha\,\mathrm{Tr}(\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}\,\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}) + \beta\left(\mathrm{Tr}(\mathbf{\Sigma}^{-1}\,\mathbf{\Omega})\right)^2 \\ &= \alpha\,\mathrm{Tr}(\mathbf{U}\,\mathbf{\Lambda}^2\,\mathbf{U}^T) + \beta\left(\mathrm{Tr}(\mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^T)\right)^2 \\ &= \alpha\,\mathrm{Tr}(\mathbf{\Lambda}^2) + \beta\left(\mathrm{Tr}(\mathbf{\Lambda})\right)^2. \end{aligned} \tag{62}$$

One can notice that $\mathrm{Tr}(\mathbf{\Lambda}^2) = \|\mathrm{diag}(\mathbf{\Lambda})\|_2^2$ and $\left(\mathrm{Tr}(\mathbf{\Lambda})\right)^2 \leq \|\mathrm{diag}(\mathbf{\Lambda})\|_1^2$, where $\mathrm{diag}(\cdot)$ returns the vector of diagonal elements of its argument, and $\|\cdot\|_2$ and $\|\cdot\|_1$ denote the L2 and L1 norms, respectively. From the Cauchy-Schwarz inequality, we have $\|\mathrm{diag}(\mathbf{\Lambda})\|_1^2 \leq M\|\mathrm{diag}(\mathbf{\Lambda})\|_2^2$. It follows that $g_{\mathbf{\Sigma}}^{ces}(\mathbf{\Omega},\mathbf{\Omega}) > 0$ if $\alpha + M\beta > 0$, completing the proof.

Proof of Theorem 4

First, the directional derivative of $g_{\mathbf{\Sigma}}^{ces}(\mathbf{\Omega}_1,\mathbf{\Omega}_2)$ in the direction $\mathbf{\Omega}_3$, where $\mathbf{\Sigma} \in \mathcal{S}_M^{++}$ and $\mathbf{\Omega}_1, \mathbf{\Omega}_2, \mathbf{\Omega}_3 \in \mathcal{S}_M$ is

$$\begin{aligned} \mathrm{D}\, g_{\mathbf{\Sigma}}^{ces}(\mathbf{\Omega}_1,\mathbf{\Omega}_2)[\mathbf{\Omega}_3] = {} & g_{\mathbf{\Sigma}}^{ces}(\mathrm{D}\,\mathbf{\Omega}_1[\mathbf{\Omega}_3],\mathbf{\Omega}_2) \\ & + g_{\mathbf{\Sigma}}^{ces}(\mathbf{\Omega}_1,\mathrm{D}\,\mathbf{\Omega}_2[\mathbf{\Omega}_3]) \\ & - \beta\,\mathrm{Tr}(\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_3\,\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_1)\,\mathrm{Tr}(\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_2) \\ & - \beta\,\mathrm{Tr}(\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_1)\,\mathrm{Tr}(\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_3\,\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_2) \\ & - \alpha\,\mathrm{Tr}(\mathbf{\Sigma}^{-1}(\mathbf{\Omega}_3\,\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_1 + \mathbf{\Omega}_1\,\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_3)\,\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_2). \end{aligned} \tag{63}$$

It then follows from the Koszul formula (equation (5.11) in [42]) that the Levi-Civita connection ∇ of $\mathbf{\Omega}_2$ in the direction $\mathbf{\Omega}_1$ on $\mathcal{S}_M^{++}$ endowed with metric (17) which is defined for all $\mathbf{\Sigma} \in \mathcal{S}_M^{++}$

$$\nabla_{\mathbf{\Omega}_1}\mathbf{\Omega}_2 = \mathrm{D}\,\mathbf{\Omega}_2[\mathbf{\Omega}_1] - \mathrm{sym}(\mathbf{\Omega}_2\,\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_1), \tag{64}$$

where $\mathrm{sym}(\cdot)$ is the operator that returns the symmetrical part of its argument. The Levi-Civita connection is the same as for the classical Riemannian metric in our case and we therefore have the same geodesics, which can be found for example in [46]. They can be characterized in two different (but equivalent) manners: the geodesics γ on $\mathcal{S}_M^{++}$ are defined for all $\mathbf{\Sigma} \in \mathcal{S}_M^{++}$ and $\mathbf{\Omega} \in \mathcal{S}_M$ as

$$\gamma(t) = \mathbf{\Sigma}^{1/2}\exp(t\,\mathbf{\Sigma}^{-1/2}\,\mathbf{\Omega}\,\mathbf{\Sigma}^{-1/2})\,\mathbf{\Sigma}^{1/2}, \tag{65}$$

where $\exp(\cdot)$ denotes the matrix exponential. Equivalently, we can define the geodesic γ between $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ in $\mathcal{S}_M^{++}$ as

$$\gamma(t) = \mathbf{\Sigma}_1^{1/2}(\mathbf{\Sigma}_1^{-1/2}\,\mathbf{\Sigma}_2\,\mathbf{\Sigma}_1^{-1/2})^t\,\mathbf{\Sigma}_1^{1/2}, \tag{66}$$

where $(\cdot)^t = \exp(t\log(\cdot))$ denotes the matrix power function defined through the matrix exponential and logarithm. Furthermore, one can check that the metric (17) is invariant by congruence, *i.e.*

$$g^{ces}_{\mathbf{U}\,\mathbf{\Sigma}\,\mathbf{U}^T}(\mathbf{U}\,\mathbf{\Omega}_1\,\mathbf{U}^T, \mathbf{U}\,\mathbf{\Omega}_2\,\mathbf{U}^T) = g^{ces}_{\mathbf{\Sigma}}(\mathbf{\Omega}_1, \mathbf{\Omega}_2), \tag{67}$$

for all $\mathbf{\Sigma} \in \mathcal{S}_M^{++}$, $\mathbf{\Omega}_1, \mathbf{\Omega}_2 \in \mathcal{S}_M$ and invertible matrix $\mathbf{U}$. Since we have the same geodesic and the congruence invariance property, the proof is completed by using the same steps given in [41] for the proof of the Riemannian distance on $\mathcal{S}_M^{++}$ equipped with the classical Riemannian metric ($\alpha = 0$ and $\beta = 1$).

Bibliography

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

[2] Steven M. Kay, *Fundamentals of statistical signal processing*, Prentice Hall PTR, 1993.

[3] John D. Gorman and Alfred O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1285–1301, 1990.

[4] Petre Stoica and Boon Chong Ng, "On the Cramér-Rao bound under parametric constraints," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 177–179, 1998.

[5] Thomas L. Marzetta, "A simple derivation of the constrained multiple parameter Cramér-Rao bound," *IEEE Transactions on Signal Processing*, vol. 41, no. 6, pp. 2247–2249, 1993.

[6] Harrie Hendriks, "A Cramér-Rao type lower bound for estimators with values in a manifold," *Journal of Multivariate Analysis*, vol. 38, no. 2, pp. 245–261, 1991.

[7] Joao Xavier and Victor Barroso, "Intrinsic distance lower bound for unbiased estimators on riemannian manifolds," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*. IEEE, 2002, vol. 2, pp. II–1141.

[8] Joao Xavier and Victor Barroso, "Intrinsic variance lower bound (IVLB): an extension of the Cramér-Rao bound to riemannian manifolds," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*. IEEE, 2005, vol. 5, pp. v–1033.

[9] Nicolas Boumal, "On intrinsic Cramér-Rao bounds for riemannian submanifolds and quotient manifolds," *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1809–1821, 2013.

[10] Silvère Bonnabel and Axel Barrau, "An intrinsic Cramér-Rao bound on Lie groups," in *Geometric Science of Information*, Frank Nielsen and Frédéric Barbaresco, Eds. 2015, pp. 664–672, Springer International Publishing.

[11] N. Boumal, A. Singer, P. A. Absil, and V. D. Blondel, "Cramér-Rao bounds for synchronization of rotations," *Information and Inference: A Journal of the IMA*, vol. 3, no. 1, pp. 1–39, March 2014.

[12] S. Bonnabel and A. Barrau, "An intrinsic Cramer-Rao bound on SO(3) for (dynamic) attitude filtering," in *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 2158–2163.

[13] Fang Kai-Tai and Zhang Yao-Ting, *Generalized multivariate analysis*, Science Press Beijing and Springer-Verlag, Berlin, 1990.

[14] Esa Ollila, Jan Eriksson, and Visa Koivunen, "Complex elliptically symmetric random variables - generation, characterization, and circularity tests," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 58–69, 2011.

[15] Esa Ollila, David E. Tyler, Visa Koivunen, and H. Vincent Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on signal processing*, vol. 60, no. 11, pp. 5597–5625, 2012.

[16] Maria Greco, Fulvio Gini, and Muralidhar Rangaswamy, "Statistical analysis of measured polarimetric clutter data at different range resolutions," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 153, no. 6, pp. 473–481, 2006.

[17] Esa Ollila, David E. Tyler, Visa Koivunen, and H. Vincent Poor, "Compound-Gaussian clutter modeling with an inverse gaussian texture distribution," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 876–879, 2012.

[18] Javier Portilla, Vasily Strela, Martin J Wainwright, and Eero P Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image processing*, vol. 12, no. 11, pp. 1338–1351, 2003.

[19] Fei Shi and Ivan W Selesnick, "An elliptically contoured exponential mixture model for wavelet based image denoising," *Applied and Computational Harmonic Analysis*, vol. 23, no. 1, pp. 131–151, 2007.

[20] Steeve Zozor and Christophe Vignat, "Some results on the denoising problem in the elliptically distributed context," *IEEE transactions on Signal Processing*, vol. 58, no. 1, pp. 134–150, 2010.

[21] Ricardo A. Maronna and Víctor J. Yohai, "Robust estimation of multivariate location and scatter," *Wiley StatsRef: Statistics Reference Online*, 1976.

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

[23] A. Wiesel, "Unified framework to regularized covariance estimation in scaled gaussian models," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 29–38, Jan 2012.

[24] E. Ollila and D. E. Tyler, "Regularized m -estimators of scatter matrix," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 6059–6070, Nov 2014.

[25] Lutz Duembgen and David E Tyler, "Geodesic convexity and regularized scatter estimators," *arXiv preprint arXiv:1607.05455*, 2016.

[26] Ann E.S. Mitchell, "The information matrix, skewness tensor and a-connections for the general multivariate elliptic distribution," *Annals of the Institute of Statistical Mathematics*, vol. 41, no. 2, pp. 289–304, 1989.

[27] Maria Greco and Fulvio Gini, "Cramér-Rao lower bounds on covariance matrix estimation for complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6401–6409, 2013.

[28] Frédéric Pascal and Alexandre Renaux, "Statistical analysis of the covariance matrix mle in k-distributed clutter," *Signal Processing*, vol. 90, no. 4, pp. 1165–1175, 2010.

[29] Olivier Besson and Yuri I. Abramovich, "On the fisher Information Matrix for multivariate elliptically contoured distributions," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1130–1133, 2013.

[30] Maia Berkane, Kevin Oden, and Peter M. Bentler, "Geodesic estimation in elliptical distributions," *Journal of Multivariate Analysis*, vol. 63, no. 1, pp. 35–46, 1997.

[31] Charles A. Micchelli and Lyle Noakes, "Rao distances," *Journal of Multivariate Analysis*, vol. 92, no. 1, pp. 97–115, 2005.

[32] Davy Paindaveine, "A canonical definition of shape," *Statistics & probability letters*, vol. 78, no. 14, pp. 2240–2247, 2008.

[33] Fulvio Gini and Maria Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, 2002.

[34] Ernesto Conte, Antonio De Maio, and Giuseppe Ricci, "Recursive estimation of the covariance matrix of a compound-gaussian process and its application to adaptive CFAR detection," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1908–1915, 2002.

[35] Frédéric Pascal, Yacine Chitour, Jean-Philippe Ovarlez, Philippe Forster, and Pascal 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, 2008.

[36] David E. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, pp. 234–251, 1987.

[37] Barrett O'neill, *Semi-Riemannian geometry with applications to relativity*, vol. 103, Academic press, 1983.

[38] William M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, vol. 120, Academic press, 1986.

[39] Isaac Chavel, *Riemannian geometry: a modern introduction*, vol. 98, Cambridge university press, 2006.

[40] Gilles Meyer, *Geometric optimization algorithms for linear regression on fixed-rank matrices*, Ph.D. thesis, Université de Liège, Belgique, 2011.

[41] Serge Lang, *Fundamentals of differential geometry*, vol. 191, Springer Science & Business Media, 2012.

[42] Pierre-Antoine Absil, Robert Mahony, and Rodolphe Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.

[43] Nicolas Boumal, *Optimization and estimation on manifolds*, Ph.D. thesis, Université catholique de Louvain, feb 2014.

[44] Axel Barrau and Silvere Bonnabel, "A note on the intrinsic Cramér-Rao bound," in *Geometric Science of Information*, pp. 377–386. Springer, 2013.

[45] Shun-ichi Amari and Hiroshi Nagaoka, *Methods of information geometry*, vol. 191, American Mathematical Soc., 2007.

[46] Rajendra Bhatia, *Positive definite matrices*, Princeton university press, 2009.