Intrinsic Cramér-Rao Bounds for Scatter and Shape Matrices Estimation in CES Distributions

Arnaud Breloy, Guillaume Ginolhac, Alexandre Renaux, Florent Bouchard

***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.**

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

I. Introduction

CRAMÉR-RAO lower bounds (CRLBs) are ubiquitous tools in statistical signal processing, as they characterize the optimum performances in terms of mean squared error (MSE) 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 CRLBs 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, so a constrained CRLB [3–5] cannot be derived for a more refined performance study. Additionally, the classical CRLB applies on the MSE (Euclidean metric), while this criterion may not be the most appropriate for characterizing the performance when parameters are living in a manifold. For example [6–9] proposed CRLBs for periodic error costs, more suited to angle estimation problems. In our context, a lower bound on the mean natural Riemannian distance can be more relevant and also reveal hidden properties of estimators.

To overcome these issues, intrinsic (i.e. manifold oriented) versions of the Cramér-Rao inequality have been established and studied in [1, 10–16]. In [1] intrinsic CRLBs are expressed in the form of a matrix inequality on the covariance of the inverse exponential map. This quantity is shown to be greater (in the matrix sense) than a matrix involving the Fisher information matrix and Riemanian curvature terms. A key property is that this inequality is valid for any chosen Riemannian metric. Thus, it allows to derive intrinsic CRLB on the distances associated to any chosen Riemannian metric. Notably, [1] obtained intrinsic CRLBs for the problem of covariance matrix estimation under the Gaussian assumption. This result provides a lower bound on the natural Riemannian distance on $\mathcal{H}_M^{++}$ (the manifold of Hermitian positive definite matrices) as well as interesting insights, e.g. the observation of a bias of the sample covariance matrix at low sample support, not exhibited by the traditional Euclidean analysis.

The aim of this work is to apply this intrinsic analysis to the class of Complex Elliptically Symmetric (CES) distributions [17–19]. These distributions provide a class that has attracted interest in the signal processing community, as it includes a large panel of well known distributions that can accurately model various physical phenomenon, such as radar clutter measurements [20, 21] or observations in image processing [22–24]. The CES distributions are in particular characterized by their scatter matrix (or its normalized counterpart, referred to as shape matrix) that is proportional to the covariance matrix if it exists. The presented results are the following: $i)$ We obtain the Fisher information metric and associated Riemannian distances induced by CES distributions in the complex case. On a side note, the latter offers generalized Riemannian distances that seems interesting for building new regularized estimators in the vein of [39–42]. $ii)$ We derive intrinsic CRLBs for the problem of scatter and shape matrices estimation under CES distribution [19]. These results extend the Euclidean CRLBs from [25–28] to various Riemannian distances, as well as a the Theorems 4 and 5 of [1] to a wider class of distributions, metrics, and for both scatter and shape matrices. An interesting note is that CRLBs for shape matrix estimation were not previously derived with practical formulation for the Euclidean metric. The proposed result allows then to draw a practical comparison of different M-estimators.

Arnaud Breloy is with LEME (EA4416), University Paris Nanterre. Guillaume Ginolhac is with LISTIC (EA3703), University Savoie Mont Blanc. Alexandre Renaux is with L2S (UMR8506), University Paris-Sud. Florent Bouchard is with GISPA-lab (UMR5216), Grenoble INP. Part of this work was supported by ANR-ASTRID MARGARITA (ANR-17-ASTR-0015).

II. CES DISTRIBUTIONS

CES distributions [17] encompass a large family of multivariate distributions. We refer the reader to the very comprehensive and detailed review on the topic in [18, 19]. A vector $\mathbf{z} \in \mathbb{C}^M$ follows a zero-mean CES distribution, denoted $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g)$ if it admits the stochastic representation $\mathbf{z} \overset{d}{=} \sqrt{\mathcal{Q}}\, \boldsymbol{\Sigma}^{1/2}\, \mathbf{u}$ where $\mathbf{u} \in \mathbb{C}^M$ follows a uniform distribution on the complex unit sphere, $\mathcal{Q} \in \mathbb{R}_+$ is non-negative real random variable of probability density function p, independent of $\mathbf{u}$, and called the second-order modular variate, and $\boldsymbol{\Sigma}^{1/2} \in \mathbb{C}^{M \times M}$ is a factorization of the scatter matrix $\boldsymbol{\Sigma}$. We focus here only on the absolute-continuous case, i.e. when $\boldsymbol{\Sigma} \in \mathcal{H}_{\mathcal{M}}^{++}$. In this case, the probability density function (pdf) of $\mathbf{z}$ is given as

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

where the function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is called the density generator and is related to the pdf of the second-order modular variate by:

$$p(\mathcal{Q}) = \delta_{M,g}^{-1} \mathcal{Q}^{M-1} g(\mathcal{Q}). \tag{2}$$

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 $\mathbf{z}$. This ambiguity is not impactful, as most of adaptive processes only require an estimate of the scatter matrix up to a scale [21]. To this end, let us define $\boldsymbol{\Sigma} = \sigma^2 \mathbf{V}$ 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 [29]. Hence $\mathbf{V}$ belongs to the manifold referred to as the special group of $\mathcal{H}_M^{++}$, denoted

$$\mathcal{SH}_M^{++} = \left\{ \mathbf{V} \in \mathcal{H}_M^{++} |\ |\mathbf{V}| = 1 \right\}. \tag{3}$$

Eventually, a simple way to redefine a CES distribution $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g)$ so that scale and shape parameters coincide is to absorb the ambiguity in the second-order modular variate as $\mathcal{Q}' \overset{d}{=} \sqrt[M]{|\boldsymbol{\Sigma}|}\mathcal{Q}$, leading to the equivalent distribution $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{V}, \tilde{g})$, where $\tilde{g}$ is set from $p(\mathcal{Q}')$ and (2).

From samples $\{\mathbf{z}_k\}_{k \in [\![1,K]\!]}$ following $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g)$ the maximum likelihood estimator of the scatter matrix is the solution of the fixed point equation [19, 30–32]

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{K} \sum_{k=1}^{K} \psi\left(\mathbf{z}_k^H \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{z}_k \right) \mathbf{z}_k \mathbf{z}_k^H, \tag{4}$$

where $\psi(t) = -g'(t)/g(t)$. Note that, in practice, the true density generator may not be known. In the robust estimation theory, an M-estimator of the scatter matrix [33, 34] refers to an estimator built using a function $\psi(t)$ that is not necessarily linked to g in (4). An important note is that M-estimators may not be consistent in scale. A practical way to remove this ambiguity is to focus on the shape matrix estimation by constructing $\hat{\mathbf{V}} = \hat{\boldsymbol{\Sigma}} / \sqrt[M]{|\hat{\boldsymbol{\Sigma}}|}$, for a given M-estimator (or MLE) of the scatter matrix $\hat{\boldsymbol{\Sigma}}$.

III. FISHER INFORMATION METRIC AND NATURAL DISTANCE INDUCED BY CES DISTRIBUTIONS

In this section we study the information geometry of the likelihood (1) on $\mathcal{H}_M^{++}$ and $\mathcal{SH}_M^{++}$. First, recall that the tangent spaces of $\mathcal{H}_M^{++}$ and $\mathcal{SH}_M^{++}$ are respectively

$$\begin{aligned} &T_{\boldsymbol{\Sigma}} \mathcal{H}_M^{++} = \mathcal{H}_M \text{ (manifold of Hermitian matrices)}, \\ &T_{\mathbf{V}} \mathcal{SH}_M^{++} = \{ \boldsymbol{\Omega} \in \mathcal{H}_M |\ \mathrm{Tr}\{\mathbf{V}^{-1} \boldsymbol{\Omega}\} = 0 \}. \end{aligned} \tag{5}$$

We have the following results:

Theorem 1 (FIM for CES) *Let $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ be in $T_{\boldsymbol{\Sigma}} \mathcal{H}_M^{++}$. The Fisher information metric associated to K i.i.d. samples following $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g)$, $\boldsymbol{\Sigma} \in \mathcal{H}_M^{++}$ is*

$$\begin{aligned} &g_{\boldsymbol{\Sigma}}^{fim}(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) = K\ g_{\boldsymbol{\Sigma}}^{ces}(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2), \ \textit{with} \\ &g_{\boldsymbol{\Sigma}}^{ces}(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) \\ &= \alpha \mathrm{Tr}\left\{ \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega}_1 \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega}_2 \right\} + \beta \mathrm{Tr}\left\{ \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega}_1 \right\} \mathrm{Tr}\left\{ \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega}_2 \right\} \end{aligned} \tag{6}$$

and

$$\alpha = 1 - \frac{\mathbb{E}\left[\mathcal{Q}^2 \phi'(\mathcal{Q}) \right]}{M(M+1)} \quad \textit{and} \quad \beta = \alpha - 1 \tag{7}$$

using $\phi(t) = -\psi(t) = g'(t)/g(t)$. Now, let $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ be in $T_{\mathbf{V}} \mathcal{SH}_M^{++}$. The Fisher information metric of the equivalent model $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{V}, \tilde{g})$, $\mathbf{V} = \boldsymbol{\Sigma} / \sigma^2 \in \mathcal{SH}_M^{++}$, is

$$\begin{aligned} &g_{\mathbf{V}}^{fim}(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) = K\ g_{\mathbf{V}}^{ces}(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2), \ \textit{with} \\ &g_{\mathbf{V}}^{ces}(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) = \tilde{\alpha} \mathrm{Tr}\left\{ \mathbf{V}^{-1} \boldsymbol{\Omega}_1 \mathbf{V}^{-1} \boldsymbol{\Omega}_2 \right\} \end{aligned} \tag{8}$$

and $\tilde{\alpha} = \alpha / \sigma^4$ (with $\sigma^2 = \sqrt[M]{|\boldsymbol{\Sigma}|}$).

Proof: The proof from [35] is extended to the complex case and $\mathcal{SH}_M^{++}$ in the supplementary materials. ■

Theorem 2 (Induced Riemannian distances) *$g_{\boldsymbol{\Sigma}}^{ces}$ in (6) is a Riemannian metric on $\mathcal{H}_M^{++}$ if and only if $\alpha > 0$ and $\alpha + M\beta > 0$. The distance induced by this metric on $\mathcal{H}_M^{++}$ is defined $\forall\, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2 \in \mathcal{H}_M^{++}$ as*

$$d_{ces}^2(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \alpha \sum_{i=1}^{M} \log^2 \lambda_i + \beta \left(\sum_{i=1}^{M} \log \lambda_i \right)^2, \tag{9}$$

where λ_i is the i^{th} eigenvalue of $\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_2$. Additionally, $g_{\mathbf{V}}^{ces}$ in (8) is a Riemannian metric on $\mathcal{SH}_M^{++}$ if and only if $\tilde{\alpha} > 0$. The distance on $\mathcal{SH}_M^{++}$ induced by this metric is defined $\forall \mathbf{V}_1, \mathbf{V}_2 \in \mathcal{SH}_M^{++}$ as

$$d_{sp-ces}^2(\mathbf{V}_1, \mathbf{V}_2) = \tilde{\alpha} \sum_{i=1}^{M} \log^2 \lambda_i, \tag{10}$$

where λ_i is the i^{th} eigenvalue of $\mathbf{V}_1^{-1} \mathbf{V}_2$.

Proof: The proof is in the supplementary material and relies on [36–38]. ■

Notice that (6) yields the classical Riemannian metric/distance on $\mathcal{H}_M^{++}$ [37] for $\alpha = 1$ and $\beta = 0$. This also corresponds to the Gaussian case in [1] since $\alpha = 1$ and $\beta = 0$ are obtained for the Gaussian density generator $g(t) = \exp(-t)$ (see [28] for the calculation of these coefficients). On the other hand, d_{sp-ces} corresponds to a scaled natural distance on $\mathcal{SH}_M^{++}$ for any underlying distribution, as the term in β vanishes in the Fisher information metric.

IV. Intrinsic CRLBs on scatter and shape matrices

In the following, we derive the intrinsic CRLBs for unbiased estimators of the scatter matrix $\hat{\mathbf{\Sigma}} \in \mathcal{H}_M^{++}$ for $\{\mathbf{z}_k\}_{k\in[\![1,K]\!]}$, i.i.d. distributed according to $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g)$. In parallel, we obtain intrinsic CRLBs for unbiased estimators of the shape matrix $\hat{\mathbf{V}}$ for the equivalent model $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{V}, \tilde{g})$ with $\mathbf{\Sigma} = \sqrt[M]{|\mathbf{\Sigma}|}\mathbf{V}$ so $\mathbf{V} \in \mathcal{SH}_M^{++}$. The terms α (resp. $\tilde{\alpha}$) and β refers to (7). The derivations rely on the steps described in [1] and appropriate construction of orthonormal basis of the tangent spaces $T_{\mathbf{\Sigma}}\mathcal{H}_M^{++}$ and $T_{\mathbf{\Sigma}}\mathcal{SH}_M^{++}$ in (5).

A. Euclidean Metric

First recall that the Euclidean metric and distance are

$$\begin{aligned} g^E(\mathbf{\Omega}_1, \mathbf{\Omega}_2) &= \mathrm{Tr}\{\mathbf{\Omega}_1\mathbf{\Omega}_2\}, \\ d_E^2(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) &= \|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_F^2. \end{aligned} \tag{11}$$

The CRLBs on d_E^2 requires the following definitions:

- $\{\mathbf{\Omega}_i^E\}_{i\in[\![1,M^2]\!]}$ denotes a basis of $T_{\mathbf{\Sigma}}\mathcal{H}_M^{++}$ in (5) that is orthonormal w.r.t. the inner product g^E in (11). In practice, we take the canonical Euclidean Basis:
  1) $\mathbf{\Omega}_{ii}^E$ is an n by n symmetric matrix whose ith diagonal element is one, zeros elsewhere
  2) $\mathbf{\Omega}_{ij}^E$ is an n by n symmetric matrix whose ijth and jith elements are both $2^{-1/2}$, zeros elsewhere.
  3) $\mathbf{\Omega}_{ij}^{h-E}$ 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 ($i < j$).

  which is re-indexed over i to lighten the notations.
- $\{\mathbf{\Omega}_i^{spE}\}_{i\in[\![1,M^2-1]\!]}$ denotes a basis of $T_{\mathbf{\Sigma}}\mathcal{SH}_M^{++}$ in (5) that is orthonormal w.r.t. the inner product g^E in (11). Remark that $T_{\mathbf{\Sigma}}\mathcal{SH}_M^{++}$ corresponds to $\mathcal{H}_M$ deprived from the line $\lambda\mathbf{\Sigma}^{-1}$, $\lambda \in \mathbb{R}$, as its orthonormal complementary is $N_{\mathbf{\Sigma}}\mathcal{SH}_M^{++} = \{\lambda\mathbf{\Sigma}^{-1}\,|\lambda\in\mathbb{R}\}$. Therefore, its orthonormal basis can be computed in practice by augmenting $\{\mathbf{\Omega}_i^E\}_{i\in[\![1,M^2]\!]}$ with the element $\mathbf{\Sigma}^{-1}$, then applying a Gram-Schmidt orthonormalization process, using the inner product g^E in (11), and starting from $\mathbf{\Sigma}^{-1}$. The output of this process is then $\left\{\gamma\mathbf{\Sigma}^{-1}, \mathbf{\Omega}_1^{spE}, \dots, \mathbf{\Omega}_{M^2-1}^{spE}, \mathbf{0}\right\}$ (with appropriate normalization γ), allowing to extract the desired basis.

Theorem 3 (Euclidean CRLBs) *The CRLBs on the distance d_E^2 for scatter and shape matrices estimation are*

$$\begin{aligned} \mathbb{E}\left[d_E^2\left(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}\right)\right] &\geq \mathrm{Tr}\left\{\mathbf{F}_E^{-1}\right\} \\ \mathbb{E}\left[d_E^2\left(\hat{\mathbf{V}}, \mathbf{V}\right)\right] &\geq \mathrm{Tr}\left\{\mathbf{F}_{spE}^{-1}\right\}, \end{aligned} \tag{12}$$

with for $i,j \in [\![1, M^2]\!]$

$$\begin{aligned} [\mathbf{F}_E]_{i,j} \;=\; & K\alpha\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}_i^E\mathbf{\Sigma}^{-1}\mathbf{\Omega}_j^E\right\} \\ & + K\beta\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}_i^E\right\}\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}_j^E\right\}, \end{aligned} \tag{13}$$

and with for $i,j \in [\![1, M^2-1]\!]$

$$[\mathbf{F}_{spE}]_{i,j} = K\tilde{\alpha}\mathrm{Tr}\left\{\mathbf{V}^{-1}\mathbf{\Omega}_i^{spE}\mathbf{V}^{-1}\mathbf{\Omega}_j^{spE}\right\}, \tag{14}$$

Proof: The Fisher information matrix entries are obtained using $g_{\mathbf{\Sigma}}^{fim}$ in (6) (resp. $g_{\mathbf{V}}^{fim}$ in (8)) and the basis $\{\mathbf{\Omega}_i^E\}$ (resp. $\{\mathbf{\Omega}_i^{spE}\}$). The result is then a direct application of Corollary 2 in [1], i.e. neglecting the Riemannian curvature terms for small errors. ■

This theorem allows to compute the Euclidean CRLB on the shape matrix in a practical way and without requiring a parameterization that ensures unit determinant. This is, to the best of our knowledge, a new result for the Euclidean distance. For the scatter matrix, we retrieve the results of [26].

B. Natural Riemannian metric

Recall that the natural Riemannian metric and associated distances on $\mathcal{H}_M^{++}$ and $\mathcal{SH}_M^{++}$ are identically defined as

$$\begin{aligned} g_{\mathbf{\Sigma}}^N(\mathbf{\Omega}_1, \mathbf{\Omega}_2) &= \mathrm{Tr}\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}_1\mathbf{\Sigma}^{-1}\mathbf{\Omega}_2\}, \\ d_N^2(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) &= \left\|\log(\mathbf{\Sigma}_1^{-1/2}\,\mathbf{\Sigma}_2\,\mathbf{\Sigma}_1^{-1/2})\right\|_F^2. \end{aligned} \tag{15}$$

The CRLBs on d_N^2 requires the following definitions:

- $\{\mathbf{\Omega}_i^N\}_{i\in 1\dots M^2}$ denotes a basis of $T_{\mathbf{\Sigma}}\mathcal{H}_M^{++}$ in (5) that is orthonormal w.r.t. the inner product $g_{\mathbf{\Sigma}}^N$ in (15). In practice, such basis can be obtained by coloring the basis of previous section as $\mathbf{\Omega}_i^N = \mathbf{\Sigma}^{1/2}\mathbf{\Omega}_i^E\mathbf{\Sigma}^{1/2}$.
- $\{\mathbf{\Omega}_i^{spN}\}_{i\in[\![1,M^2-1]\!]}$ denotes a basis of $T_{\mathbf{\Sigma}}\mathcal{SH}_M^{++}$ in (5) that is orthonormal w.r.t. the inner product $g_{\mathbf{\Sigma}}^N$ in (15). In practice, such basis can be obtained by the same process as for $\{\mathbf{\Omega}_i^{spE}\}$, but using the inner product $g_{\mathbf{\Sigma}}^N$ in (15) to perform the orthonormalization process. Note that the initial basis should however be augmented with $\mathbf{\Sigma}$ here, since the orthonormal complementary of $T_{\mathbf{\Sigma}}\mathcal{SH}_M^{++}$ w.r.t. $g_{\mathbf{\Sigma}}^N$ is $N_{\mathbf{\Sigma}}\mathcal{SH}_M^{++} = \{\lambda\,\mathbf{\Sigma}\,|\lambda\in\mathbb{R}\}$.

Theorem 4 (Natural Riemmanian CRLBs) *The CRLBs on the distance d_N^2 for scatter and shape matrices estimation are*

$$\begin{aligned} \mathbb{E}\left[d_N^2\left(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}\right)\right] &\geq \frac{M^2-1}{K\alpha} + (K(\alpha + M\beta))^{-1} \\ \mathbb{E}\left[d_N^2\left(\hat{\mathbf{V}}, \mathbf{V}\right)\right] &\geq \frac{M^2-1}{K\tilde{\alpha}} \end{aligned} \tag{16}$$

Proof: For the scatter matrix we plug the basis $\{\mathbf{\Omega}_i^N\} = \{\mathbf{\Sigma}^{1/2}\mathbf{\Omega}_i^E\mathbf{\Sigma}^{1/2}\}$ into $g_{\mathbf{\Sigma}}^{fim}$ in (6). The entries of the Fisher information matrix are then

$$g_{\mathbf{\Sigma}}^{fim}\left(\mathbf{\Omega}_i^N, \mathbf{\Omega}_j^N\right) = K\alpha\mathrm{Tr}\left\{\mathbf{\Omega}_i^E\mathbf{\Omega}_j^E\right\} + K\beta\mathrm{Tr}\left\{\mathbf{\Omega}_i^E\right\}\mathrm{Tr}\left\{\mathbf{\Omega}_j^E\right\}.$$

With proper ordering of $\{\mathbf{\Omega}_i^E\}$ this matrix is obtained as

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

which reads $\mathbf{F}_N = K\alpha\mathbf{I} + KM\beta\mathbf{v}_{fim}\mathbf{v}_{fim}^H$ with $\mathbf{v}_{fim} = {}^1/\sqrt{M}\left[\,\mathbf{1}_M \mid \mathbf{0}_{M(M-1)}\,\right]$. Hence the eigenvalues of $\mathbf{F}_N^{-1}$ can be easily identified as $K^{-1}\left[(\alpha+M\beta)^{-1}, \alpha^{-1}, \dots, \alpha^{-1}\right]$ and summed to obtain its trace. For the shape matrix, we plug the basis $\{\mathbf{\Omega}_i^{spN}\}$ into (8). This gives the entries of the Fisher Information Matrix as $[\mathbf{F}_{spN}]_{i,j} = K\tilde{\alpha}\delta_{i,j},\ \forall i,j \in [\![1, M^2-1]\!]$, thanks to the orthonormality of $\{\mathbf{\Omega}_i^{spN}\}$ w.r.t. $g_{\mathbf{\Sigma}}^{nat}$. The Fisher information matrix is therefore $\mathbf{F}_{spN} = K\tilde{\alpha}\mathbf{I}_{M^2-1}$ whose the trace of inverse reads directly. The results are then applications of Corollary 2 in [1]. ■

C. CES-Fisher Information Metric

Recall that the CES-Fisher information metric and associated distance are given in (6) and (9) respectively. We denote $\{\mathbf{\Omega}_i^{ces}\}_{i\in 1\dots M^2}$, a basis of $T_{\mathbf{\Sigma}}\mathcal{H}_M^{++}$ that is orthonormal w.r.t. to the metric (inner product) $g_{\mathbf{\Sigma}}^{ces}$. Closed-form expressions of this basis are not needed for the developments, but it can be constructed numerically in practice.

Remark: Note that, from Theorem 2, the CES-Fisher (10) distance on $\mathcal{SH}_M^{++}$ corresponds to the natural Riemannian distance (15) on $\mathcal{SH}_M^{++}$ scaled by α. Hence, regarding to the shape estimation, the Theorem 4 holds for both Natural and CES-Fisher distances up this scale factor in definition of the estimation error.

Theorem 5 (CES-Fisher CRLB) *The CRLBs on the distance d_{ces}^2 for scatter matrix estimation is*

$$\mathbb{E}\left[d_{ces}^2\left(\hat{\mathbf{\Sigma}},\mathbf{\Sigma}\right)\right] \geq \frac{M^2}{K} \tag{17}$$

Proof: The Fisher information matrix entries are obtained by plugging the basis $\{\mathbf{\Omega}_i^{ces}\}$ in $g_{\mathbf{\Sigma}}^{fim}$ in (6). Notice that $g_{\mathbf{\Sigma}}^{fim} = K g_{\mathbf{\Sigma}}^{ces}$, 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 the Corollary 2 in [1]. ■

V. Simulations

Previous results are illustrated for the multivariate Student t-distribution with $d \in \mathbb{N}^*$ degree of freedom (see [19] for details). We have $\mathbf{z} \sim \mathcal{CES}(\mathbf{0},\mathbf{\Sigma},g_d)$ with $g_d(t) = (1+d^{-1}t)^{-(d+M)}$, hence $\phi(t) = -(d+M)/(d+t)$ and $\alpha = (d+M)/(d+M+1)$ in (7). 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 coincide. We consider the following estimators of the scatter: $a)$ SCM, defined as $\hat{\mathbf{\Sigma}}_{SCM} = K^{-1}\sum_{k=1}^{K}\mathbf{z}_k\mathbf{z}_k^H$, $b)$ MLE,defined in (4) using $\psi(t) = -\phi(t)$, $c)$ Mismatched MLE, defined as MLE except that we use $d = 10$ in ψ regardless of the underlying distribution, $d)$ Tyler's M-estimator, defined in (4) 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 by re-normalization. To empirically validate the obtained results, we compare the performance of the different estimators to the corresponding CRLB in two settings: $d = 100$ (close to Gaussian case) and $d = 3$. Figure 1 displays the performance w.r.t. d_N^2 and d_{ces}^2 (d_E^2 is omitted since it has been studied in [26]) in terms of scatter matrix estimation. Figure 2 displays the performance w.r.t. d_E^2 and d_N^2 (proportional to d_{ces}^2) in terms of shape matrix estimation.

In Figure 1, for $d = 100$, $\alpha \simeq 1$ and $\beta \simeq 0$, so $g_{\mathbf{\Sigma}}^{nat}$ and $g_{\mathbf{\Sigma}}^{ces}$ generate almost identical distances and corresponding CRLBs, 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. For $d = 3$, we note that the SCM and the mismatched MLE (due to the bias induced on its scale) have poor performance as expected.

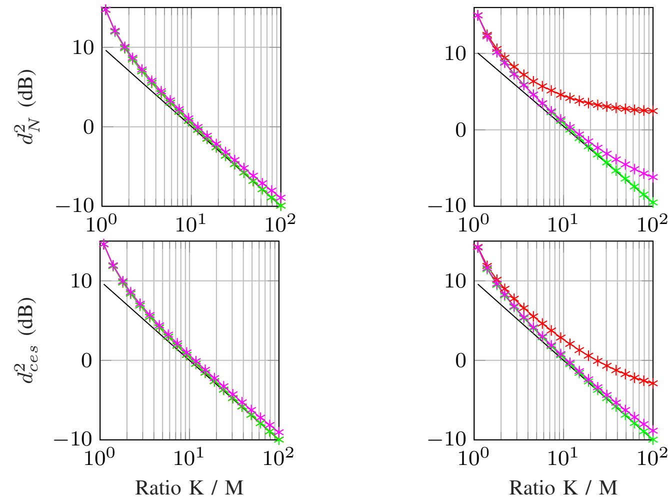

Fig. 1. Natural Riemannian (top) and CES-Fisher (bottom) CRLB and corresponding mean squared distance of scatter matrix estimators for t-distribution versus K/M. Legend: CRLB on the considered distance (black), SCM (red), MLE (green), Mismatched MLE (magenta). $M = 10$, and $d = 100$ (left) or $d = 3$ (right).

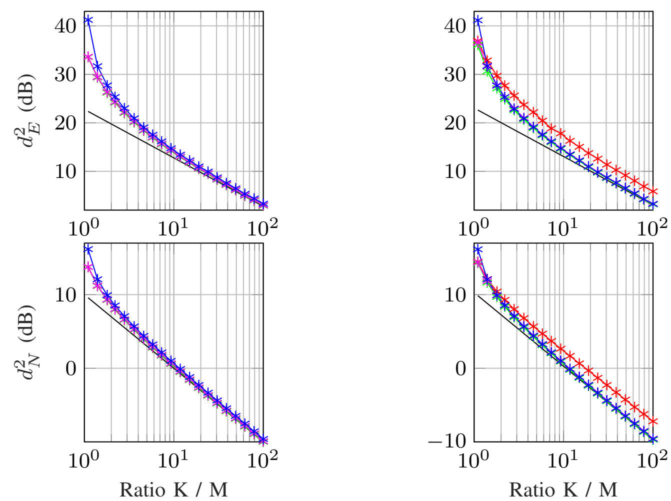

Fig. 2. Euclidean (top) and Natural Riemannian (bottom) CRLB and corresponding mean squared distance of shape matrix estimators for t-distribution versus K/M. Legend: CRLB on the considered distance (black), SCM (red), MLE (green), Mismatched MLE (magenta), Tyler (blue). $M = 10$, and $d = 100$ (left) or $d = 3$ (right).

Conversely, Figure 2 illustrates that intrinsic CRLBs on shape 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. Indeed, such comparison is 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 close to Gaussian ($d = 3$).

References

[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] Tirza Routtenberg and Joseph Tabrikian, "Non-Bayesian Periodic Cramér-Rao bound," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 1019–1032, 2013.

[7] Tirza Routtenberg and Joseph Tabrikian, "Cyclic Barankin Bounds for Non-Bayesian Periodic Parameter estimation," *IEEE Transactions on Signal Processing*, vol. 62, no. 13, pp. 3321–3336, 2014.

[8] Tirza Routtenberg and Joseph Tabrikian, "Bayesian Parameter Estimation using Periodic Cost Functions," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1229–1240, 2012.

[9] Eyal Nitzan, Tirza Routtenberg and Joseph Tabrikian, "A New Class of Bayesian Cyclic Bounds for Periodic Parameter Estimation," *IEEE Transactions on Signal Processing*, vol. 64, no. 1, pp. 229–243, 2016.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Nicolas Boumal, Amit Singer, Pierre-Antoine Absil, and Vincent 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.

[16] Silvère Bonnabel and Axel 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.

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

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

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

[30] 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.

[31] 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.

[32] 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.

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

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

[35] 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.

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

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

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

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

[40] 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.

[41] 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.

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

VI. Supplementary materials

A. Proof of Theorem 1

For $\mathbf{\Omega} \in T_{\mathbf{\Sigma}}\mathcal{H}_M^{++}$, the Fisher Information Metric is obtained according to Theorem 1 of [1] as

$$g_{\mathbf{\Sigma}}^{fim}(\mathbf{\Omega},\mathbf{\Omega}) = -\mathbb{E}\left[\left. \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{L}\left(\{\mathbf{z}_k\}|\mathbf{\Sigma}+t\partial\mathbf{\Omega},g\right)\right|_{t=0} \right]. \quad (18)$$

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

$$\mathcal{L}\left(\{\mathbf{z}_k\}|\mathbf{\Sigma},g\right) = \sum_{k=1}^{K}\log\left(g\left(\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right)\right) - K\log|\mathbf{\Sigma}|, \quad (19)$$

where $\mathbf{Z}_k = \mathbf{z}_k\mathbf{z}_k^H$. Through the second-order Taylor expansions around $\mathbf{\Sigma}$, one can obtain

$$\begin{aligned}\left.\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{L}\left(\{\mathbf{z}_k\}|\mathbf{\Sigma}+t\mathbf{\Omega}\right)\right|_{t=0} &= K\mathrm{Tr}\left\{\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\right\} \\ &+ 2\sum_{k=1}^{K}\mathrm{Tr}\left\{\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\mathbf{Z}_k\mathbf{\Sigma}^{-1}\right\}\phi\left(Tr\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right) \\ &+ \sum_{k=1}^{K}\mathrm{Tr}^2\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\phi'\left(Tr\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right).\end{aligned} \quad (20)$$

In order to compute the expectations, we recall 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:

$$\begin{aligned}&\mathbb{E}\left[\mathrm{Tr}\left\{\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\mathbf{Z}_k\mathbf{\Sigma}^{-1}\right\}\phi\left(\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right)\right] \\ &\mathbb{E}\left[\mathrm{Tr}\left\{\mathbf{\Sigma}^{H/2}\mathbf{\Sigma}^{-1}\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\mathbf{\Sigma}^{1/2}\mathbf{u}_k\mathbf{u}_k^H\right\}\right]\mathbb{E}\left[\mathcal{Q}_k\phi\left(\mathcal{Q}_k\right)\right] \\ &= -\mathrm{Tr}\left\{\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\right\},\end{aligned} \quad (21)$$

where we used $\mathbb{E}\left[\mathbf{u}_k\mathbf{u}_k^H\right] = \mathbf{I}_M/M$ (since $\mathbf{u}_k \sim \mathcal{U}(\mathbb{CS}^M)$), and (2) to obtain the result $\mathbb{E}\left[\mathcal{Q}_k\phi\left(\mathcal{Q}_k\right)\right] = -M$. The second expectation is obtained by the same method as

$$\begin{aligned}&\mathbb{E}\left[\mathrm{Tr}^2\left\{\mathbf{\Sigma}^{-1}\mathbf{\Omega}\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\phi'\left(\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{Z}_k\right\}\right)\right] \\ &= \frac{\mathbb{E}\left[\mathcal{Q}_k^2\phi'\left(\mathcal{Q}_k\right)\right]}{M(M+1)}\left(\mathrm{Tr}^2\left\{\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right\} + \mathrm{Tr}\left\{\left(\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^2\right\}\right),\end{aligned} \quad (22)$$

where we used the relation from [28], giving

$$\mathbb{E}\left[\left(\mathbf{u}_k^H\mathbf{B}\mathbf{u}_k\right)^2\right] = \left(\mathrm{Tr}\left\{\mathbf{B}^2\right\} + \mathrm{Tr}^2\left\{\mathbf{B}\right\}\right)/\left(M(M+1)\right), \quad (23)$$

for an arbitrary constant matrix $\mathbf{B}$ and $\mathbf{u}_k \sim \mathcal{U}\left(\mathbb{CS}^M\right)$. Eventually, by plugging (21) and (22) into (18) and (20), the Fisher Information Metric is given as:

$$g_{\mathbf{\Sigma}}^{fim}(\mathbf{\Omega},\mathbf{\Omega}) = K\alpha\mathrm{Tr}\left\{\left(\mathbf{\Sigma}^{-1}\mathbf{\Omega}\right)^2\right\} + K\beta\mathrm{Tr}^2\left\{\mathbf{\Omega}\mathbf{\Sigma}^{-1}\right\}, \quad (24)$$

with coefficients α and β defined in (7). Notice that the dependency on k in $\mathcal{Q}$ is omitted since these parameters are assumed to be i.i.d.. Also, some manipulations with $\phi'(t) = g''(t)/g(t) - \phi^2(t)$ and (2) allow to show that

$$M(M+1) - \mathbb{E}\left[\mathcal{Q}^2\phi'\left(\mathcal{Q}\right)\right] = \mathbb{E}\left[\mathcal{Q}^2\phi^2\left(\mathcal{Q}\right)\right], \quad (25)$$

which is consistent with the coefficients obtained in the parametric case of [28]. To obtain the metric $g_{\mathbf{\Sigma}}^{fim}(\mathbf{\Omega}_1,\mathbf{\Omega}_2)$ use a standard polarization formula which, after some expansions and simplifications leads to (6) and (7).

To derive the metric (8) we first remark that

$$g_{\mathbf{\Sigma}}^{ces}(\mathbf{\Omega}_1,\mathbf{\Omega}_2) = \tilde{\alpha}\mathrm{Tr}\left\{\mathbf{V}^{-1}\mathbf{\Omega}_1\mathbf{V}^{-1}\mathbf{\Omega}_2\right\} + \tilde{\beta}\mathrm{Tr}\left\{\mathbf{V}^{-1}\mathbf{\Omega}_1\right\}\mathrm{Tr}\left\{\mathbf{V}^{-1}\mathbf{\Omega}_2\right\}.$$

With such expression, $g_{\mathbf{V}}^{fim}$ can be identified to $g_{\mathbf{\Sigma}}^{fim}$ on the restricted set $\mathcal{SH}_M^{++}$, as they describe the same underlying distribution. Note that this is also consistent with the change of variable $\mathbf{\Sigma} = \sigma^2\mathbf{V}$ and Theorem 3 of [1]. Second, we also notice that the terms $\mathrm{Tr}\left\{\mathbf{V}^{-1}\mathbf{\Omega}\right\}$ are equal to 0 when $\mathbf{\Omega} \in T_{\mathbf{V}}\mathcal{SH}_M^{++}$ which cancels all the terms in β in (8).

B. Proof of Theorem 2

It is readily checked that for all $\mathbf{\Sigma} \in \mathcal{H}_M^{++}$ the function (6) is symmetric and bilinear. It remains to determine whether it is positive-definite. Let $\mathbf{\Sigma} \in \mathcal{H}_M^{++}$ and $\mathbf{\Omega} \in \mathcal{H}_M$, 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

$$\begin{aligned}g_{\mathbf{\Sigma}}^{ces}(\mathbf{\Omega},\mathbf{\Omega}) &= \alpha\,\mathrm{Tr}(\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}\,\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}) + \beta\left(\mathrm{Tr}(\mathbf{\Sigma}^{-1}\,\mathbf{\Omega})\right)^2 \\ &= \alpha\,\mathrm{Tr}(\mathbf{U}\,\mathbf{\Lambda}^2\,\mathbf{U}^T) + \beta\left(\mathrm{Tr}(\mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^T)\right)^2 \\ &= \alpha\,\mathrm{Tr}(\mathbf{\Lambda}^2) + \beta\left(\mathrm{Tr}(\mathbf{\Lambda})\right)^2.\end{aligned} \quad (26)$$

One can notice that $\mathrm{Tr}(\mathbf{\Lambda}^2) = \|\mathrm{diag}(\mathbf{\Lambda})\|_2^2$ and $(\mathrm{Tr}(\mathbf{\Lambda}))^2 \leq \|\mathrm{diag}(\mathbf{\Lambda})\|_1^2$, 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 $\alpha + M\beta > 0$. Now, the directional derivative of $g_{\mathbf{\Sigma}}^{ces}(\mathbf{\Omega}_1,\mathbf{\Omega}_2)$ in the direction $\mathbf{\Omega}_3$, where $\mathbf{\Sigma} \in \mathcal{H}_M^{++}$ and $\mathbf{\Omega}_1,\mathbf{\Omega}_2,\mathbf{\Omega}_3 \in T_{\mathbf{\Sigma}}\mathcal{H}_M^{++}$ is

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

It then follows from the Koszul formula (equation (5.11) in [38]) that the Levi-Civita connection ∇ of $\mathbf{\Omega}_2$ in the direction $\mathbf{\Omega}_1$ on $\mathcal{H}_M^{++}$ endowed with metric (6) which is defined for all $\mathbf{\Sigma} \in \mathcal{H}_M^{++}$

$$\nabla_{\mathbf{\Omega}_1}\mathbf{\Omega}_2 = \mathrm{D}\,\mathbf{\Omega}_2[\mathbf{\Omega}_1] - \mathrm{sym}(\mathbf{\Omega}_2\,\mathbf{\Sigma}^{-1}\,\mathbf{\Omega}_1), \quad (27)$$

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, i.e. γ on $\mathcal{H}_M^{++}$, defined for all $\mathbf{\Sigma} \in \mathcal{H}_M^{++}$ and $\mathbf{\Omega} \in T_{\mathbf{\Sigma}}\mathcal{H}_M^{++}$ as

$$\gamma(t) = \mathbf{\Sigma}^{1/2}\exp(t\,\mathbf{\Sigma}^{-1/2}\,\mathbf{\Omega}\,\mathbf{\Sigma}^{-1/2})\,\mathbf{\Sigma}^{1/2}, \quad (28)$$

where $\exp(\cdot)$ denotes the matrix exponential. Furthermore, one can check that the metric (6) is invariant by congruence, *i.e.*

$$g_{\mathbf{U}\,\mathbf{\Sigma}\,\mathbf{U}^T}^{ces}(\mathbf{U}\,\mathbf{\Omega}_1\,\mathbf{U}^T,\mathbf{U}\,\mathbf{\Omega}_2\,\mathbf{U}^T) = g_{\mathbf{\Sigma}}^{ces}(\mathbf{\Omega}_1,\mathbf{\Omega}_2), \quad (29)$$

for all $\mathbf{\Sigma} \in \mathcal{H}_M^{++}$, $\mathbf{\Omega}_1,\mathbf{\Omega}_2 \in T_{\mathbf{\Sigma}}\mathcal{H}_M^{++}$ 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 [36] for the proof of the Riemannian distance on $\mathcal{H}_M^{++}$ equipped with the classical Riemannian metric (where $\alpha = 1$ and $\beta = 0$). Therefore, we can show that the distance on $\mathcal{H}_M^{++}$ is equal to (9). The distance (10) on $\mathcal{SH}_M^{++}$ can be obtained by following the same steps.