# A Riemannian Framework for Low-Rank Structured Elliptical Models

Florent Bouchard, Arnaud Breloy, Guillaume Ginolhac, Alexandre Renaux, Frédéric Pascal

# A Riemannian Framework for Low-Rank Structured Elliptical Models

Florent Bouchard, Arnaud Breloy, *Member, IEEE*, Guillaume Ginolhac, *Senior Member, IEEE*, Alexandre Renaux, *Member, IEEE*, Frédéric Pascal, *Senior Member, IEEE*

***Abstract*—This paper proposes an original Riemmanian geometry for low-rank structured elliptical models, *i.e.*, when samples are elliptically distributed with a covariance matrix that has a low-rank plus identity structure. The considered geometry is the one induced by the product of the Stiefel manifold and the manifold of Hermitian positive definite matrices, quotiented by the unitary group. One of the main contribution is to consider an original Riemannian metric, leading to new representations of tangent spaces and geodesics. From this geometry, we derive a new Riemannian optimization framework for robust covariance estimation, which is leveraged to minimize the popular Tyler's cost function on the considered quotient manifold. We also obtain a new divergence function, which is exploited to define a geometrical error measure on the quotient, and the corresponding intrinsic Cramér-Rao lower bound is derived. Thanks to the structure of the chosen parametrization, we further consider the subspace estimation error on the Grassmann manifold and provide its intrinsic Cramér-Rao lower bound. Our theoretical results are illustrated on some numerical experiments, showing the interest of the proposed optimization framework and that performance bounds can be reached.**

***Index Terms*—Riemannian geometry, elliptical distributions, robust estimation, covariance matrix, low-rank structure, Cramér-Rao bounds,**

## I. Introduction

Complex elliptically symmetric distributions offer a general family of statistical models that encompasses most of standard multivariate distributions, including the Gaussian one, as well as many heavy-tailed distributions, such as multivariate Student $t$-, and $K$- distributions (cf. [1] for a review on this topic). These models have been leveraged successfully in numerous applications thanks to their good empirical fit to datasets, *e.g.*, in image processing [2]–[4] or array processing [5], [6]. On top of that, elliptical models have also attracted a lot of interest, as they allow robust estimation processes to be derived. For example, $M$-estimators [7], [8], defined as generalized maximum likelihood estimators of elliptical models, have been shown to be robust to model mismatches and contaminated data (outliers) [1]. While alleviating robustness issues, the development of estimation algorithms under elliptical models is still challenged by "small $n$ large $p$" problems (where $n$ and $p$ respectively stand for the sample size and the dimension).

In several applications, one can rightfully assume that the relevant information lies in a low dimensional subspace. This is reflected by a low-rank structure of the covariance matrix, often referred to as spiked model [9]. This idea plays a central role in principal component analysis [10], [11], subspace recovery [12], and related dimension reduction algorithms. In array processing, low-rank models are also at the core of subspace methods [13], low-rank adaptive filters [14] and detectors [15]. These structures are also involved in financial time series analysis [16] (where they are also referred to as factor models). Estimation processes in such low-rank models have been well studied for Gaussian distributions [10], [17]. Unfortunately, the results obtained in this case cannot be trivially transposed to elliptical distributions. For example, low-rank structured counterparts of $M$-estimators are not expressed in closed form, nor directly tractable. Additionally, ultimate statistical performance characterization is not obvious in this context, due to constraints/ambiguities on the parameters space.

This paper proposes to leverage tools from Riemannian geometry in order to answer the previous questions with a unified view. The Riemannian standpoint was adopted in [18] to derive intrinsic (*i.e.*, manifold oriented) Cramér-Rao lower bounds, then applied to study both unstructured and low-rank Gaussian models. This leads to interesting results and insights, such as performance bounds for various Riemannian distances, and the characterization of a bias of the sample covariance matrix at low sample support, not exhibited by the traditional Euclidean analysis. The Riemannian geometry of the manifold of Hermitian positive definite matrices has also been recently used to study unstructured elliptical models. It notably revealed hidden (geodesic) convexity properties of elliptical distribution's likelihood functions [19], and allowed to derive new regularization-based estimation algorithms [20]–[22]. Studying low-rank elliptical models require to turn to the manifold of Hermitian positive semi-definite matrices of fixed rank $k$ ($k < p$), which has, to the best of our knowledge, not been proposed in this context. The contributions associated to the proposed framework for low-rank elliptical models follow three main axes, summed up below.

### A. Geometry for low-rank structured elliptical models

The statistical parameter of the considered low-rank model for complex elliptically symmetric distributions lives in the

Florent Bouchard and Guillaume Ginolhac are with LISTIC (EA3703), University Savoie Mont Blanc, France (e-mails: florent.bouchard@univ-smb.fr, guillaume.ginolhac@univ-smb.fr). Arnaud Breloy is with LEME (EA4416), University Paris Nanterre, France (e-mail: abreloy@parisnanterre.fr). Alexandre Renaux is with Laboratoire des signaux et systèmes (UMR8506), University Paris-Sud, France (e-mail: alexandre.renaux@u-psud.fr). Frederic Pascal is with Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes (UMR8506), 91190, Gif-sur-Yvette, France (e-mail: frederic.pascal@l2s.centralesupelec.fr). This work was supported by ANR-ASTRID MARGARITA (ANR-17-ASTR-0015).

manifold $\mathcal{H}_{p,k}^{+}$ of $p \times p$ Hermitian positive semi-definite matrices of rank $k$. This manifold has recently attracted much attention and several geometries have been proposed for it; see *e.g.*, [23]–[27]. In this work, we consider the geometry induced by the quotient $(\mathrm{St}_{p,k} \times \mathcal{H}_k^{++})/\mathcal{U}_k$, *i.e.,* the product manifold of the complex Stiefel manifold $\mathrm{St}_{p,k}$ of $p \times k$ orthogonal matrices (with $p > k$) and the manifold $\mathcal{H}_k^{++}$ of $k \times k$ Hermitian positive definite matrices, quotiented by the unitary group $\mathcal{U}_k$. This geometry has already been studied in the context of low-rank matrices in [23], [25]. It is of particular interest in our context because the principal subspace of the covariance matrix is directly obtained from this parametrization and a divergence function, which can be exploited to measure estimation errors, is available in closed form [23].

Our framework differs from the works [23], [25] as we propose a new Riemannian metric on the product $\mathrm{St}_{p,k} \times \mathcal{H}_k^{++}$: the part on $\mathrm{St}_{p,k}$ is the so-called canonical metric on Stiefel [28] while the part on $\mathcal{H}_k^{++}$ is a general form of the affine invariant metric which corresponds to the Fisher information metric of elliptical distributions on $\mathcal{H}_k^{++}$ [29]. As a direct consequence, the representations of tangent spaces of the quotient $(\mathrm{St}_{p,k} \times \mathcal{H}_k^{++})/\mathcal{U}_k$, geodesics, Riemannian gradient and Hessian used for optimization are original in this context. We also introduce a retraction, which corresponds to a second order approximation of the geodesics. Moreover, we derive a new divergence function and its associated geometry on the quotient, which is inspired by the one of [23].

### B. Algorithms for robust low-rank covariance matrix estimation

Covariance matrix estimation is a crucial step in many machine learning and signal processing algorithms. In elliptical models, $M$-estimators [7], [8] offer a robust alternative to the traditional sample covariance matrix. These estimators appear as generalized maximum likelihood estimators and ensure good asymptotic properties [1], [30], [31]. Nevertheless, $M$-estimators do not account for the low-rank structure. A natural solution to this issue is to directly derive an estimator as the minimizer of a robust cost function under a low-rank structure constraint. This approach has been proposed in [32, Sec. V.A.], where a majorization-minimization algorithm is proposed to minimize Tyler's cost function according to this structure. However, the tractability of this estimator is an open question at low sample support (*cf.* assumption 2 in [32]). Notably, the majorization-minimization algorithm can present convergence issues in some practical case where $n$ is close to or smaller than $p$.

To address this issue, we propose to use the Riemannian optimization framework [33]: the proposed geometry for the the quotient $(\mathrm{St}_{p,k} \times \mathcal{H}_k^{++})/\mathcal{U}_k$ indeed offers the possibility to apply a large panel of generic first and second order optimization algorithms on manifolds, such as gradient descent, conjugate gradient, BFGS, trust region, Newton, *etc.* (*cf.* [33] for details). More specifically for robust covariance matrix estimation, we propose an estimator formulated as the minimizer of a counterpart of Tyler's cost function defined directly on $(\mathrm{St}_{p,k} \times \mathcal{H}_k^{++})/\mathcal{U}_k$. We then focus on two algorithms for solving the introduced problem: one based on Riemannian gradient descent (first order method), the other based on Riemannian trust region (second order method). In terms of estimation accuracy, our numerical experiments show that the Riemannian trust region based algorithm is similar to [32, algorithm 5]. Interestingly, these experiments also show that the Riemannian gradient descent based method can still reach good performance when the other methods diverge at insufficient sample support.

### C. Statistical performance analysis in low-rank elliptical models

Cramér-Rao lower bounds are ubiquitous tools to characterize the optimum performances in terms of mean squared error that can be achieved for a given parametric estimation problem [34]. In the context of elliptical distributions, Cramér-Rao lower bounds can be obtained using the general results of [35], and have been studied for covariance/shape estimation in [36], [37]. However, the low-rank models involve constraints and ambiguities on the parameters space, which does not allow for simple/practical derivations, even using the so-called constrained Cramér-Rao lower bounds [38]–[40]. Additionally, the classical inequality applies on the mean squared error (Euclidean metric), while this criterion may not be the most appropriate for characterizing the performance when parameters are living in a manifold. To overcome these issues, intrinsic (i.e. Riemannian manifold oriented) versions of the Cramér-Rao inequality have been established and studied in [18], [41]. Interestingly, the obtained inequalities are valid for any chosen Riemannian metric. Thus, these results allows to derive performance bounds on any Riemannian distance used as error measure for the estimation of a parameter living in a manifold[1].

Leveraging this framework for covariance matrix estimation in low-rank elliptical models, we first derive the performance bound on the Riemannian distance related to the considered metric on the quotient $(\mathrm{St}_{p,k} \times \mathcal{H}_k^{++})/\mathcal{U}_k$ (total error measurement). However, this distance does not admit a closed-form expression, which is why the proposed divergence appears as a practical alternative. For this divergence, we then derive an alternative intrinsic Cramér-Rao bound that goes beyond the standard Riemannian geometry framework in [18], [41]. Finally, we focus on the Riemannian distance on the Grassmann manifold $\mathcal{G}_{p,k}$ [28] and derive the corresponding performance bound for principal subspace estimation. Some numerical experiments then illustrate that the obtained performance bounds can be reached by the proposed algorithms. These contributions therefore generalize the ones of [18] on low-rank Gaussian models to wider classes of distributions and performance measures, and the results of [29] to low-rank models.

[1] In some situations, the standard Euclidean Cramér-Rao bound can be recovered as a special case: this occurs when the Euclidean distance actually corresponds to the Riemannian one when the manifold is endowed with the euclidean metric. However, this is will not be the case for parameter space considered in this paper.

## II. Review of Riemannian geometry, optimization and intrinsic Cramér-Rao bounds

Before dealing with low-rank structured elliptical models, general recalls on Riemannian geometry, optimization and intrinsic Cramér-Rao bounds are provided. All the tools presented in this section, which can be found in [18], [33], [41], [42], are later employed to tackle robust estimation and performance analysis in the context of low-rank covariance matrices. Section II-A deals with smooth manifolds while section II-B focuses on quotient manifolds.

### A. Smooth manifold

*a) Geometry:* A *smooth manifold* $\mathcal{M}$ is a space which is locally diffeomorphic to a vector space and which admits a differential structure, *i.e.*, every point $\theta \in \mathcal{M}$ possesses a *tangent space* $T_\theta\mathcal{M}$, whose elements are called tangent vectors and generalize the concept of directional derivatives. Such $\mathcal{M}$ is turned into a Riemannian manifold by endowing it with a *Riemannian metric* $\langle\cdot,\cdot\rangle_\cdot$, which is a smoothly varying inner product on every tangent space $T_\theta\mathcal{M}$.

One often needs to handle vector fields on $\mathcal{M}$, *i.e.*, functions associating one tangent vector in $T_\theta\mathcal{M}$ to each point $\theta \in \mathcal{M}$. Directional derivatives of vector fields are generalized with affine connections. On a Riemannian manifold $\mathcal{M}$, a specific affine connection plays a particular role: the *Levi-Civita connection* $\nabla_\cdot\cdot$, which is characterized by Koszul formula. The Levi-Civita connection allows to define *geodesics* $\gamma : I \subseteq \mathbb{R} \to \mathcal{M}$, which generalize the concept of straight lines in $\mathcal{M}$. They are indeed curves with zero acceleration, *i.e.*, $\nabla_{\dot\gamma(t)}\dot\gamma(t) = 0_{\gamma(t)}$, where $0_{\gamma(t)}$ is the zero element of $T_{\gamma(t)}\mathcal{M}$. Geodesics $\gamma$ only depend on the choice of initial point $\gamma(0) = \theta \in \mathcal{M}$ and initial direction $\dot\gamma(0) = \xi \in T_\theta\mathcal{M}$. Given $\theta \in \mathcal{M}$, the *Riemannian exponential mapping* is the mapping from $T_\theta\mathcal{M}$ onto $\mathcal{M}$ such that $\exp_\theta(\xi) = \gamma(1)$, where $\gamma$ is the geodesic such that $\gamma(0) = \theta$ and $\dot\gamma(0) = \xi$. Its inverse, the *Riemannian logarithm mapping*, can also be defined: Given $\theta \in \mathcal{M}$, it is the mapping from $\mathcal{M}$ onto $T_\theta\mathcal{M}$ such that $\log_\theta(\vartheta) = \xi$, where $\exp_\theta(\xi) = \vartheta$. The *Riemannian distance* $\delta$ on $\mathcal{M}$ is then obtained through

$$\delta^2(\theta, \vartheta) = \|\log_\theta(\vartheta)\|_\theta^2 = \langle\log_\theta(\vartheta), \log_\theta(\vartheta)\rangle_\theta.$$

*b) Optimization:* Given an objective function $f : \mathcal{M} \to \mathbb{R}$, the *Riemannian gradient* of $f$ at $\theta \in \mathcal{M}$ is defined through the Riemannian metric as the unique tangent vector in $T_\theta\mathcal{M}$ such that, for all $\xi \in T_\theta\mathcal{M}$,

$$\langle \mathrm{grad}_{\mathcal{M}} f(\theta), \xi\rangle_\theta = \mathrm{D}\, f(\theta)[\xi],$$

where $\mathrm{D}\, f(\theta)[\xi]$ is the directional derivative of $f$ at $\theta$ in direction $\xi$. The *Riemannian Hessian* of $f$ at $\theta \in \mathcal{M}$ in direction $\xi \in T_\theta\mathcal{M}$ is defined as $\mathrm{Hess}_{\mathcal{M}} f(\theta)[\xi] = \nabla_\xi \,\mathrm{grad}\, f(\theta)$.

A descent direction $\xi \in T_\theta\mathcal{M}$ of $f$ at $\theta$ can be obtained from the Riemannian gradient and Hessian of $f$. A new point on the manifold is then achieved by a *retraction* $R_\theta : T_\theta\mathcal{M} \to \mathcal{M}$, which is a mapping such that $R_\theta(0_\theta) = \theta$ and for all $\xi \in T_\theta\mathcal{M}$, $\mathrm{D}\, R_\theta(0_\theta)[\xi] = \xi$. A Riemannian manifold admits a natural retraction: the Riemannian exponential mapping. However, for numerical complexity and stability reasons, alternative solutions are often prefered [33].

These tools are enough to employ a large panel of first and second order Riemannian optimization algorithms such as gradient descent, Newton, trust region, *etc.* For instance, given iterate $\theta_i$, the Riemannian gradient descent algorithm yields

$$\theta_{i+1} = R_{\theta_i}(-t_i \,\mathrm{grad}_{\mathcal{M}} f(\theta_i)),$$

where $t_i$ is the stepsize, which can be computed with a line search.

*c) Intrinsic Cramér-Rao bound:* Let $\widehat\theta \in \mathcal{M}$ an unbiased estimator of the true parameters $\theta \in \mathcal{M}$ of some distribution with log-likelihood $L : \mathcal{M} \to \mathbb{R}$. Let $\{e_i\}$ an orthonormal basis of $T_\theta\mathcal{M}$ according to the metric $\langle\cdot,\cdot\rangle_\theta$. Following the Riemannian framework of [18], the estimation error associated to the metric $\langle\cdot,\cdot\rangle_\theta$ is contained in the vector $\boldsymbol{x_\theta}$, whose $i^{\text{th}}$ element is $(\boldsymbol{x}_\theta)_i = \langle\log_\theta(\widehat\theta), e_i\rangle_\theta$ (which reads as the standard error vector $\boldsymbol{x}_\theta = \theta - \widehat\theta$ in the Euclidean setting), and the covariance matrix of this error is denoted $\boldsymbol{C}_\theta = \boldsymbol{x}_\theta\boldsymbol{x}_\theta^T$. The $ij^{\text{th}}$ element of the corresponding Fisher information matrix $\boldsymbol{F}_\theta$ is $(\boldsymbol{F}_\theta)_{ij} = g_\theta^{\mathcal{M}}(e_i, e_j)$, where $g_\theta^{\mathcal{M}}(\xi,\eta) = -\mathbb{E}[\mathrm{D}^2\, L(\theta)[\xi,\eta]]$ is the Fisher information metric of the considered distribution. The intrinsic Cramér-Rao lower bound [18] on $\mathcal{M}$ is then given as

$$\mathbb{E}[\boldsymbol{C}_\theta] \succeq \boldsymbol{F}_\theta^{-1} + \text{ curvature terms.}$$

For small errors, the curvature terms can be neglected (this will be the case in this paper), and taking the trace of the above inequality yields the lower bound

$$\mathbb{E}[\delta^2(\theta, \widehat\theta)] \succeq \mathrm{tr}(\boldsymbol{F}_\theta^{-1}),$$

Where $\delta$ is the Riemannian distance associated to $\langle\cdot,\cdot\rangle_\theta$. Also notice that this expression reduces to the standard Cramér-Rao lower bound $\mathbb{E}[\|\theta - \widehat\theta\|_F^2] \succeq \mathrm{tr}(\boldsymbol{F}_\theta^{-1})$ in the Euclidean case.

### B. Quotient manifold

*a) Geometry:* A *quotient manifold* $\mathcal{M}$ of a smooth manifold $\overline{\mathcal{M}}$ is quite abstract. Its elements are indeed equivalence classes on $\overline{\mathcal{M}}$. To handle elements of $\mathcal{M}$, the usual technique is to exploit the canonical projection $\pi : \overline{\mathcal{M}} \to \mathcal{M}$, which associates $\theta = \pi(\bar\theta) \in \mathcal{M}$ to all $\bar\theta \in \overline{\mathcal{M}}$. The equivalence class of $\bar\theta$ is obtained on $\overline{\mathcal{M}}$ by $\pi^{-1}(\pi(\bar\theta))$. Every element $\theta \in \mathcal{M}$ can be represented by any element $\bar\theta \in \overline{\mathcal{M}}$ such that $\theta = \pi(\bar\theta)$. More generally, all geometrical tools of $\mathcal{M}$ can be characterized through such representations.

A Riemannian metric on $\mathcal{M}$ is defined through a metric $\langle\cdot,\cdot\rangle_\cdot$ on $\overline{\mathcal{M}}$ that is invariant along the equivalence classes $\pi^{-1}(\pi(\bar\theta))$. The tangent space $T_\theta\mathcal{M}$ at $\theta = \pi(\bar\theta) \in \mathcal{M}$ can be represented by a well chosen subspace of $T_{\bar\theta}\overline{\mathcal{M}}$. The subspace of $T_{\bar\theta}\overline{\mathcal{M}}$ inducing a move along the equivalence class $\pi^{-1}(\pi(\bar\theta))$ is the *vertical space* $\mathcal{V}_{\bar\theta} = T_{\bar\theta}\pi^{-1}(\pi(\bar\theta))$. Any complementary space to $\mathcal{V}_{\bar\theta}$ in $T_{\bar\theta}\overline{\mathcal{M}}$, called a *horizontal space*, provides unique represenatives of tangent vectors in $T_\theta\mathcal{M}$. One horizontal space is particularly interesting: the orthogonal complement to $\mathcal{V}_{\bar\theta}$ according to $\langle\cdot,\cdot\rangle_\cdot$, denoted $\mathcal{H}_{\bar\theta}$. Indeed, it turns $\pi$ into a Riemannian submsersion, *i.e.*, it is the adequate horizontal space to describe the Riemannian geometry of $\mathcal{M}$.

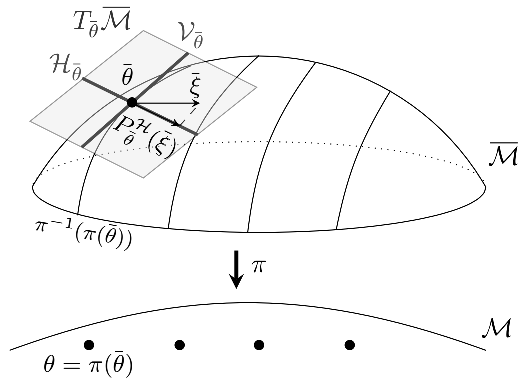


Fig. 1: Illustration of the quotient manifold $\mathcal{M}$ of manifold $\overline{\mathcal{M}}$. The tangent space $T_{\bar{\theta}}\overline{\mathcal{M}}$ can be decomposed into two complementary subspaces: the vertical space $\mathcal{V}_{\bar{\theta}} = T_{\bar{\theta}}\pi^{-1}(\pi(\bar{\theta}))$ and the horizontal space $\mathcal{H}_{\bar{\theta}}$, which provides proper representatives of tangent vectors in $T_\theta\mathcal{M}$ at $\theta = \pi(\bar{\theta})$. The orthogonal projection map $P^{\mathcal{H}}_{\bar{\theta}}$ allows to project $\bar{\xi} \in T_{\bar{\theta}}\overline{\mathcal{M}}$ onto $\mathcal{H}_{\bar{\theta}}$.

Further notice that one can define the orthogonal projection map $P^{\mathcal{H}}_{\bar{\theta}} : T_{\bar{\theta}}\overline{\mathcal{M}} \to \mathcal{H}_{\bar{\theta}}$. An illustration is provided in figure 1.

Let $\overline{\nabla}$ the Levi-Civita connection on $\overline{\mathcal{M}}$. Let $\theta = \pi(\bar{\theta}) \in \mathcal{M}$ and $\xi, \eta \in T_\theta\mathcal{M}$ represented by $\bar{\xi}, \bar{\eta} \in \mathcal{H}_{\bar{\theta}}$. The representative in $\mathcal{H}_{\bar{\theta}}$ of the Levi-Civita connection $\nabla_\xi\, \eta$ is $\overline{\nabla_\xi\, \eta} = P^{\mathcal{H}}_{\bar{\theta}}(\overline{\nabla}_{\bar{\xi}}\, \bar{\eta})$. Finally, geodesics on $\mathcal{M}$ can be defined through those on $\overline{\mathcal{M}}$. Indeed, if $\bar{\gamma}$ is a geodesic in $\overline{\mathcal{M}}$ that stays horizontal, *i.e.*, its derivative $\dot{\bar{\gamma}}(t)$ belongs to $\mathcal{H}_{\gamma(t)}$, then $\pi \circ \bar{\gamma}$ is a geodesic on $\mathcal{M}$.

*b) Optimization:* Let $\bar{f}$ an objective function on $\overline{\mathcal{M}}$. The function $\bar{f}$ induces a function $f$ on $\mathcal{M}$ if it is invariant along every equivalence class $\pi^{-1}(\pi(\bar{\theta}))$ in $\overline{\mathcal{M}}$. One then has $\bar{f} = f \circ \pi$. It follows that the gradient of $f$ at $\theta = \pi(\bar{\theta})$ is simply represented by $\mathrm{grad}_{\overline{\mathcal{M}}}\, \bar{f}(\bar{\theta})$, which belongs to $\mathcal{H}_{\bar{\theta}}$. Moreover, the representative of the Riemannian Hessian of $f$ at $\theta = \pi(\bar{\theta})$ in direction $\xi \in T_\theta\mathcal{M}$ represented by $\bar{\xi} \in \mathcal{H}_{\bar{\theta}}$ is

$$\overline{\mathrm{Hess}_{\mathcal{M}}\, f(\theta)[\xi]} = P^{\mathcal{H}}_{\bar{\theta}}(\mathrm{Hess}_{\overline{\mathcal{M}}}\, \bar{f}(\bar{\theta})[\bar{\xi}]).$$

Finally, a retraction $\overline{R}$ on $\overline{\mathcal{M}}$ induces a retraction $R$ on $\mathcal{M}$ if it is invariant along equivalence classes. Let $\bar{\theta}, \bar{\vartheta} \in \pi^{-1}(\theta)$ and $\xi \in T_\theta\mathcal{M}$ represented by $\bar{\xi} \in \mathcal{H}_{\bar{\theta}}$ and $\bar{\zeta} \in \mathcal{H}_{\bar{\vartheta}}$, respectively. A retraction $\overline{R}$ is invariant if $\pi(\overline{R}_{\bar{\theta}}(\bar{\xi})) = \pi(\overline{R}_{\bar{\vartheta}}(\bar{\zeta}))$.

*c) Intrinsic Cramér-Rao bound:* Let $\widehat{\theta} \in \mathcal{M}$ an unbiased estimator of parameters $\theta = \pi(\bar{\theta}) \in \mathcal{M}$ of some distribution with log-likelihood $L : \mathcal{M} \to \mathbb{R}$ induced by $\bar{L} : \overline{\mathcal{M}} \to \mathbb{R}$. Let $\{h_i\}$ an orthonormal basis of $\mathcal{H}_{\bar{\theta}}$. The error matrix is $\boldsymbol{C}^{\mathcal{H}}_{\bar{\theta}} = \boldsymbol{x}^{\mathcal{H}}_{\bar{\theta}}(\boldsymbol{x}^{\mathcal{H}}_{\bar{\theta}})^T$ with $(\boldsymbol{x}^{\mathcal{H}}_{\bar{\theta}})_i = \langle \overline{\log_\theta(\widehat{\theta})}, h_i \rangle_{\bar{\theta}}$, where $\overline{\log_\theta(\widehat{\theta})}$ is the representative of $\log_\theta(\widehat{\theta})$ in $\mathcal{H}_{\bar{\theta}}$. The $ij^{th}$ element of the corresponding Fisher information matrix is $(\boldsymbol{F}^{\mathcal{H}}_{\bar{\theta}})_{ij} = g^{\overline{\mathcal{M}}}_{\bar{\theta}}(h_i, h_j)$, where $g^{\overline{\mathcal{M}}}$ is the Fisher information metric on $\overline{\mathcal{M}}$ associated with $\bar{L}$. The lower bound is then $\mathbb{E}[\boldsymbol{C}^{\mathcal{H}}_{\bar{\theta}}] \succeq (\boldsymbol{F}^{\mathcal{H}}_{\bar{\theta}})^{-1}$.

Sometimes, it is easier or more convenient to work on $T_{\bar{\theta}}\overline{\mathcal{M}}$. Let $\{\bar{e}_i\}$ an orthonormal basis of $T_{\bar{\theta}}\overline{\mathcal{M}}$. One can define the error matrix as $\boldsymbol{C}_{\bar{\theta}} = \boldsymbol{x}_{\bar{\theta}}\boldsymbol{x}^T_{\bar{\theta}}$ with $(\boldsymbol{x}_{\bar{\theta}})_i = \langle \overline{\log_\theta(\widehat{\theta})}, \bar{e}_i \rangle_{\bar{\theta}}$. In this case, the $ij^{th}$ element of the corresponding Fisher information matrix is $(\boldsymbol{F}_{\bar{\theta}})_{ij} = g^{\overline{\mathcal{M}}}_{\bar{\theta}}(\bar{e}_i, \bar{e}_j)$ and the lower bound is $\mathbb{E}[\boldsymbol{C}_{\bar{\theta}}] \succeq \boldsymbol{F}^{\dagger}_{\bar{\theta}}$, where $\cdot^{\dagger}$ denotes the Moore-Penrose pseudo-inverse. Taking the traces of both inequalities yield the same lower bound:

$$\begin{aligned}\mathbb{E}[\delta^2(\theta, \widehat{\theta})] = \mathbb{E}[\mathrm{tr}(\boldsymbol{C}^{\mathcal{H}}_{\bar{\theta}})] = \mathbb{E}[\mathrm{tr}(\boldsymbol{C}_{\bar{\theta}})] \\ \geq \mathrm{tr}((\boldsymbol{F}^{\mathcal{H}}_{\bar{\theta}})^{-1}) = \mathrm{tr}(\boldsymbol{F}^{\dagger}_{\bar{\theta}}).\end{aligned}$$

## III. Model

### A. Complex elliptically symmetric distributions and robust covariance estimation

Complex elliptically symmetric distributions [43] represent a large family of multivariate distributions that encompasses, for example, Gaussian, $K$-, Student $t$-, and Weibull distributions. A detailed review on the topic can be found in [1]. The probability density function (pdf) associated with the random variable $\boldsymbol{x} \in \mathbb{C}^p$ following a centered complex elliptically symmetric distribution is, up to a normalization factor,

$$f^{++}_g(\boldsymbol{x}|\boldsymbol{R}) = \det(\boldsymbol{R})^{-1} g(\boldsymbol{x}^H \boldsymbol{R}^{-1} \boldsymbol{x}), \tag{1}$$

where $\det$ denotes the determinant operator, $\boldsymbol{R} \in \mathcal{H}^{++}_p$ is the covariance matrix[2] and $g : \mathbb{R}^+ \to \mathbb{R}^+$ is the so-called density generator of the distribution.

The negative log-likelihood function associated with $n$ independent and identically distributed samples $\{\boldsymbol{x}_i\}_{i=1}^n$ of the random variable $\boldsymbol{x}$ is

$$L^{++}_g(\boldsymbol{R}) = n \log\det(\boldsymbol{R}) - \sum_{i=1}^{n} \log(g(\boldsymbol{x}_i^H \boldsymbol{R}^{-1} \boldsymbol{x}_i)). \tag{2}$$

Given the density generator $g$ and $n$ observations $\{\boldsymbol{x}_i\}$, an estimator $\widehat{\boldsymbol{R}}$ of the true covariance matrix $\boldsymbol{R}$ can be obtained by solving the maximum likelihood optimization problem

$$\widehat{\boldsymbol{R}} = \underset{\boldsymbol{R}}{\mathrm{argmin}} \quad L^{++}_g(\boldsymbol{R}).$$

Unfortunately, the true density generator $g$ is often unknown in practice. To overcome this issue, a solution provided by the robust estimation theory is to compute an $M$-estimator [7]. A popular choice is Tyler's $M$-estimator [8], [44], which is motivated by its "distribution-free" properties among the whole familly of CES, its good asymptotic performance [44], and robustness properties. Given $\{\boldsymbol{x}_i\}$, the corresponding cost function to be minimized corresponds to $g(t) = 1/t$ and is defined as

$$L^{++}_{\mathrm{T}}(\boldsymbol{R}) = p \sum_{i=1}^{n} \log(\boldsymbol{x}_i^H \boldsymbol{R}^{-1} \boldsymbol{x}_i) + n \log\det(\boldsymbol{R}). \tag{3}$$

On $\mathcal{H}^{++}_p$, this cost function is efficiently minimized with a fixed-point algorithm [44]. Additional assumptions on the structure of the covariance $\boldsymbol{R}$ can also be made; see *e.g.*, [32] for various possibilities. In this work, we are interested in the low-rank covariance structure (cf. [32, section V.A]).

[2] We adopt here a slight abuse of denomination. Rigorously, $\boldsymbol{R}$ can be referred to as the *scatter* matrix, which is proportional to the covariance matrix of $\boldsymbol{x}$, i.e., $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^H] \propto \boldsymbol{R}$ when this quantity exists.

### B. Low-rank covariance model and parameter space

Following from the probabilistic principal component analysis framework [10], the low-rank covariance model (also known as spiked model [9] or factor model [16]) refers to the structure

$$\boldsymbol{R} = \boldsymbol{I}_p + \boldsymbol{H}, \tag{4}$$

where $\boldsymbol{I}_p$ denotes the $p$-dimensional identity matrix and $\boldsymbol{H}$ is a $p \times p$ Hermitian positive semi-definite matrix of rank $k$. As done in many works, we assume the rank $k$ to be known (*e.g.*, from prior physical considerations [45]) or pre-estimated (*e.g.*, from model order selection techniques [46]).

**Remark 1.** *In general, one can consider a scaling of the white noise variance*[3]*, i.e.,* $\boldsymbol{R} = \sigma^2\boldsymbol{I}_p + \boldsymbol{H}$ *with* $\sigma^2 \in \mathbb{R}^+$*. In the considered framework, this parameter can be absorbed by the inherent scaling ambiguities of the elliptical distributions. Indeed a change in the scale of the scatter matrix from* $\boldsymbol{R}$ *to* $\tilde{\boldsymbol{R}} = \boldsymbol{R}/\sigma^2$ *can be absorbed in the density generator using* $\tilde{g}(t) = g(\sigma^{-2}t)$*, which yields an equivalent model. Hence, from* $\boldsymbol{H} \in \mathcal{H}_{p,k}^+$ *and a given pair* $(g, \sigma^2\boldsymbol{I}_p + \boldsymbol{H})$*, it is always possible to recast an equivalent distribution using* $(\tilde{g}, \boldsymbol{I}_p + \tilde{\boldsymbol{H}})$ *with still* $\tilde{\boldsymbol{H}} \in \mathcal{H}_{p,k}^+$*, meaning that omitting* $\sigma^2$ *can be done without loss of generality from a modeling point of view. However, from the estimation point of view, the density generator* $g$ *and/or the scaling* $\sigma^2$ *are generally unknown in practice. This motivates the derivation of robust and scale-free estimation processes (as discussed in Section V-B).*

The parameter $\boldsymbol{H}$ in (4) lives in the manifold $\mathcal{H}_{p,k}^+$ of $p \times p$ Hermitian positive semi-definite matrices of rank $k$. As explained in the introduction, several geometries have been proposed for this manifold. In this work, we consider the geometry resulting from the decomposition

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H, \text{ with } (\boldsymbol{U}, \boldsymbol{\Sigma}) \in \overline{\mathcal{M}}_{p,k} = (\mathrm{St}_{p,k} \times \mathcal{H}_k^{++}), \tag{5}$$

which is directly related to the singular value decomposition of $\boldsymbol{H}$. This parametrization is particularly interesting when it comes to the signal subspace estimation as the latter is simply obtained from the component $\boldsymbol{U}$.

Let $\bar{\varphi} : \overline{\mathcal{M}}_{p,k} \to \mathcal{H}_{p,k}^+$ be the smooth mapping defined, for $(\boldsymbol{U}, \boldsymbol{\Sigma}) \in \overline{\mathcal{M}}_{p,k}$, as

$$\bar{\varphi}(\boldsymbol{U}, \boldsymbol{\Sigma}) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H. \tag{6}$$

Since every $\boldsymbol{H} \in \mathcal{H}_{p,k}^+$ admits a decomposition of the form (5), the mapping $\bar{\varphi}$ is surjective. However, it is not injective as the considered decomposition is not unique: given any $\boldsymbol{O} \in \mathcal{U}_k$, one has $\boldsymbol{H} = \bar{\varphi}(\boldsymbol{U}, \boldsymbol{\Sigma}) = \bar{\varphi}(\boldsymbol{U}\boldsymbol{O}, \boldsymbol{O}^H\boldsymbol{\Sigma}\boldsymbol{O})$. As done in [23], [25], to account for the action of the unitary matrices, we define the quotient manifold

$$\mathcal{M}_{p,k} = \{\pi(\boldsymbol{U}, \boldsymbol{\Sigma}) : (\boldsymbol{U}, \boldsymbol{\Sigma}) \in \overline{\mathcal{M}}_{p,k}\}, \tag{7}$$

where the equivalence class $\pi(\boldsymbol{U}, \boldsymbol{\Sigma})$ is

$$\pi(\boldsymbol{U}, \boldsymbol{\Sigma}) = \{(\boldsymbol{U}\boldsymbol{O}, \boldsymbol{O}^H\boldsymbol{\Sigma}\boldsymbol{O}) : \boldsymbol{O} \in \mathcal{U}_k\}. \tag{8}$$

As shown in [23], [25], it follows that the function $\varphi$ on $\mathcal{M}_{p,k}$ induced by $\bar{\varphi}$ on $\overline{\mathcal{M}}_{p,k}$, *i.e.*, such that $\bar{\varphi} = \varphi \circ \pi$, is an isomorphism from $\mathcal{M}_{p,k}$ onto $\mathcal{H}_{p,k}^+$. Thus, the geometry of $\mathcal{M}_{p,k}$ can be exploited to treat problems defined on $\mathcal{H}_{p,k}^+$. In particular, the pdf on $\mathcal{M}_{p,k}$ of a random variable $\boldsymbol{x}$ following a zero-mean complex elliptically symmetric distribution with covariance matrix admitting structure (5) is, for all $\theta = \pi(\boldsymbol{U}, \boldsymbol{\Sigma}) \in \mathcal{M}_{p,k}$,

$$f_g(\boldsymbol{x}|\theta) = f_g^{++}(\boldsymbol{x}|\boldsymbol{I}_p + \varphi(\theta)), \tag{9}$$

where $f_g^{++}$ is defined in (1). Similarly, the cost function on $\mathcal{M}_{p,k}$ of the Tyler's $M$-estimator is defined, for all $\theta = \pi(\boldsymbol{U}, \boldsymbol{\Sigma}) \in \mathcal{M}_{p,k}$, as

$$L_{\mathrm{T}}(\theta) = L_{\mathrm{T}}^{++}(\boldsymbol{I}_p + \varphi(\theta)), \tag{10}$$

where $L_{\mathrm{T}}^{++}$ is defined in (3).

## IV. Riemannian geometry of Hermitian positive semi-definite matrices of fixed rank

As explained in section II, the geometry of the quotient $\mathcal{M}_{p,k}$ can be described by exploiting $\pi : \overline{\mathcal{M}}_{p,k} \to \mathcal{M}_{p,k}$ defined in (8), which allows to work with representatives in $\overline{\mathcal{M}}_{p,k}$ of the geometrical objects of $\mathcal{M}_{p,k}$. For our study, the following elements are provided: a Riemannian metric on $\overline{\mathcal{M}}_{p,k}$, invariant along equivalence classes; the corresponding horizontal spaces, properly representing tangent spaces of $\mathcal{M}_{p,k}$; the Levi-Civita connection and the associated geodesics. Unfortunately, the Riemannian logarithm map and distance remain unknown.

In the following, $\bar{\theta} = (\boldsymbol{U}, \boldsymbol{\Sigma})$, $\bar{\xi} = (\boldsymbol{\xi}_{\boldsymbol{U}}, \boldsymbol{\xi}_{\boldsymbol{\Sigma}})$, $\bar{\eta} = (\boldsymbol{\eta}_{\boldsymbol{U}}, \boldsymbol{\eta}_{\boldsymbol{\Sigma}})$ and $\overline{Z} = (\boldsymbol{Z}_{\boldsymbol{U}}, \boldsymbol{Z}_{\boldsymbol{\Sigma}})$. First recall that

$$T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k} = \{\bar{\xi} \in \mathbb{C}^{p\times k} \times \mathcal{H}_k : \boldsymbol{U}^H\boldsymbol{\xi}_{\boldsymbol{U}} + \boldsymbol{\xi}_{\boldsymbol{U}}^H\boldsymbol{U} = \boldsymbol{0}\}. \tag{11}$$

We equip $\overline{\mathcal{M}}_{p,k}$ with the Riemannian metric of definition 1. The part of this metric that concerns $\boldsymbol{U}$ is the so-called canonical metric on Stiefel [28][4], which is obtained by treating $\mathrm{St}_{p,k}$ as the quotient $\mathcal{U}_p/\mathcal{U}_{p-k}$. The one that concerns $\boldsymbol{\Sigma}$ corresponds to a class of affine invariant metrics on $\mathcal{H}_k^{++}$ that are of interest when dealing with elliptical distributions as they are related to the Fisher information metric [29][5].

**Definition 1** (Riemannian metric)**.** *We define the Riemannian metric* $\langle\cdot,\cdot\rangle_{\cdot}$ *on* $\overline{\mathcal{M}}_{p,k}$ *by*

$$\begin{aligned}\langle\bar{\xi}, \bar{\eta}\rangle_{\bar{\theta}} &= \mathfrak{Re}(\mathrm{tr}(\boldsymbol{\xi}_{\boldsymbol{U}}^H(\boldsymbol{I}_p - \frac{1}{2}\boldsymbol{U}\boldsymbol{U}^H)\boldsymbol{\eta}_{\boldsymbol{U}})) \\ &+ \alpha\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}_{\boldsymbol{\Sigma}}) + \beta\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_{\boldsymbol{\Sigma}})\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}_{\boldsymbol{\Sigma}}),\end{aligned} \tag{12}$$

*where* $\alpha > 0$ *and* $\beta > -\frac{\alpha}{k}$.

It is readily checked that the metric (12) is invariant along the equivalence classes (8), *i.e.*, for all $\boldsymbol{O} \in \mathcal{U}_k$

$$\langle\bar{\xi}, \bar{\eta}\rangle_{\bar{\theta}} = \langle\phi_{\boldsymbol{O}}(\bar{\xi}), \phi_{\boldsymbol{O}}(\bar{\eta})\rangle_{\phi_{\boldsymbol{O}}(\bar{\theta})},$$

[3]One might also be interested in the general model $\boldsymbol{R} = \boldsymbol{R}_0 + \boldsymbol{H}$, where the identity $\boldsymbol{I}_p$ is replaced by any (known) matrix $\boldsymbol{R}_0 \in \mathcal{H}_p^{++}$, as done in [18]. It is equivalent to our model as it suffices to whiten the random variable $\boldsymbol{x}$ with $\boldsymbol{R}_0^{-1/2}$ in order to obtain (4).

[4]This metric is advantageous as compared to the Euclidean metric because resulting geodesics admit simpler formulas [28].

[5]For example, the Fisher information metric on $\mathcal{H}_k^{++}$ for the Gaussian distribution is obtained with $\alpha = 1$ and $\beta = 0$.

where $\phi_{\boldsymbol{O}}(\overline{Z}) = (\boldsymbol{Z_U}\boldsymbol{O}, \boldsymbol{O}^H\boldsymbol{Z_\Sigma}\boldsymbol{O})$. Thus, metric (12) induces a Riemannian metric on the quotient $\mathcal{M}_{p,k}$. Furthermore, the orthogonal projection map according to (12) from $\mathbb{C}^{p\times k}\times\mathbb{C}^{k\times k}$ onto $T_{\bar\theta}\overline{\mathcal{M}}_{p,k}$ is

$$P_{\bar\theta}(\overline{Z}) = (\boldsymbol{Z_U} - \boldsymbol{U}\,\mathrm{herm}(\boldsymbol{U}^H\boldsymbol{Z_U}), \mathrm{herm}(\boldsymbol{Z_\Sigma})), \qquad (13)$$

where herm returns the Hermitian part of its argument.

In order to obtain the horizontal space at $\bar\theta$, one first needs to define the vertical space, which, as shown in [23], [25], is in our case given by

$$\mathcal{V}_{\bar\theta} = \{(\boldsymbol{U\Omega}, \boldsymbol{\Sigma\Omega}-\boldsymbol{\Omega\Sigma}) : \boldsymbol{\Omega}\in\mathcal{H}_k^\perp\},$$

where $\mathcal{H}_k^\perp$ denotes the space of skew-Hermitian matrices. The horizontal space along with the orthogonal projection map from $T_{\bar\theta}\overline{\mathcal{M}}_{p,k}$ onto $\mathcal{H}_{\bar\theta}$ are given in proposition 1.

**Proposition 1.** *The horizontal space $\mathcal{H}_{\bar\theta}$ at $\bar\theta\in\overline{\mathcal{M}}_{p,k}$ is*

$$\mathcal{H}_{\bar\theta} = \{\bar\xi\in T_{\bar\theta}\overline{\mathcal{M}}_{p,k} : \boldsymbol{U}^H\boldsymbol{\xi_U} = 2\alpha(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma} - \boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1})\}.$$

*The orthogonal projection map $P^{\mathcal{H}}_{\bar\theta}$ according to (12) from $T_{\bar\theta}\overline{\mathcal{M}}_{p,k}$ onto $\mathcal{H}_{\bar\theta}$ is given by*

$$P^{\mathcal{H}}_{\bar\theta}(\bar\xi) = (\boldsymbol{\xi_U} - \boldsymbol{U\Omega}, \boldsymbol{\xi_\Sigma} + \boldsymbol{\Omega\Sigma} - \boldsymbol{\Sigma\Omega}),$$

*where $\boldsymbol{\Omega}\in\mathcal{H}_k^\perp$ is the unique solution to*

$$(1-4\alpha)\boldsymbol{\Omega} + 2\alpha(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega\Sigma} + \boldsymbol{\Sigma\Omega\Sigma}^{-1}) = \boldsymbol{U}^H\boldsymbol{\xi_U} + 2\alpha(\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}).$$

*Proof:* By definition, $\bar\xi\in\mathcal{H}_{\bar\theta}$ if and only if, for all $\boldsymbol{\Omega}\in\mathcal{H}_k^\perp$, $\langle\bar\xi, (\boldsymbol{U\Omega}, \boldsymbol{\Sigma\Omega}-\boldsymbol{\Omega\Sigma})\rangle_{\bar\theta} = 0$. Since $\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma\Omega}-\boldsymbol{\Omega\Sigma})) = 0$, we have

$$\begin{aligned}\langle\bar\xi, (\boldsymbol{U\Omega}, \boldsymbol{\Sigma\Omega}-\boldsymbol{\Omega\Sigma})\rangle_{\bar\theta} &= \tfrac{1}{2}\Re\mathfrak{e}(\mathrm{tr}(\boldsymbol{\xi_U}^H\boldsymbol{U\Omega})) + \alpha\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma\Omega}-\boldsymbol{\Omega\Sigma}))\\ &= \tfrac{1}{2}\Re\mathfrak{e}(\mathrm{tr}((\boldsymbol{\xi_U}^H\boldsymbol{U} + 2\alpha(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma} - \boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1}))\boldsymbol{\Omega})).\end{aligned}$$

We thus need $\Re\mathfrak{e}(\mathrm{tr}((\boldsymbol{\xi_U}^H\boldsymbol{U}+2\alpha(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}-\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1}))\boldsymbol{\Omega})) = 0$. This is true for all $\boldsymbol{\Omega}\in\mathcal{H}_k^\perp$ if and only if $\boldsymbol{\xi_U}^H\boldsymbol{U}+2\alpha(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}-\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1})$ is Hermitian. This translates into $\boldsymbol{U}^H\boldsymbol{\xi_U} - \boldsymbol{\xi_U}^H\boldsymbol{U} = 4\alpha(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}-\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1})$. From (11), we have $\boldsymbol{U}^H\boldsymbol{\xi_U}+\boldsymbol{\xi_U}^H\boldsymbol{U} = \boldsymbol{0}$, leading to the result.

Regarding $P^{\mathcal{H}}$, it has the proposed form by definition. The matrix $\boldsymbol{\Omega}\in\mathcal{H}_k^\perp$ must be chosen in order to have $P^{\mathcal{H}}_{\bar\theta}(\bar\xi)\in\mathcal{H}_{\bar\theta}$. Basic calculations yield the proposed equation. It remains to show that the solution exists and is unique. This equation can be vectorized as

$$((1-4\alpha)\boldsymbol{I}_{k^2} + 2\alpha(\boldsymbol{\Sigma}^{-T}\otimes\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^T\otimes\boldsymbol{\Sigma}^{-1}))\,\mathrm{vec}(\boldsymbol{\Omega}) = \mathrm{vec}(\boldsymbol{U}^H\boldsymbol{\xi_U} + 2\alpha(\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma})).$$

Showing that $(1-4\alpha)\boldsymbol{I}_{p^2} + 2\alpha(\boldsymbol{\Sigma}^{-T}\otimes\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^T\otimes\boldsymbol{\Sigma}^{-1})$ is positive definite is enough to conclude. In order to do so, consider the eigenvalue decomposition $\boldsymbol{\Sigma} = \boldsymbol{V\Lambda V}^H$. We have

$$((1-4\alpha)\boldsymbol{I}_{k^2} + 2\alpha(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}^{-1})) = (\overline{\boldsymbol{V}}\otimes\boldsymbol{V})((1-4\alpha)\boldsymbol{I}_{k^2}+2\alpha(\boldsymbol{\Lambda}^{-1}\otimes\boldsymbol{\Lambda}+\boldsymbol{\Lambda}\otimes\boldsymbol{\Lambda}^{-1}))(\overline{\boldsymbol{V}}\otimes\boldsymbol{V})^H,$$

where $\overline{\boldsymbol{V}}$ is the conjugate of $\boldsymbol{V}$. As $\boldsymbol{V}$ is unitary, $\overline{\boldsymbol{V}}$ and $\overline{\boldsymbol{V}}\otimes\boldsymbol{V}$ are also unitary. $((1-4\alpha)\boldsymbol{I}_{p^2} + 2\alpha(\boldsymbol{\Lambda}^{-1}\otimes\boldsymbol{\Lambda} + \boldsymbol{\Lambda}\otimes\boldsymbol{\Lambda}^{-1}))$ is diagonal and its elements are $1-4\alpha+2\alpha(\frac{\lambda_i}{\lambda_j}+\frac{\lambda_j}{\lambda_i})$, where $\lambda_i$ is the $i^{\text{th}}$ diagonal element of $\boldsymbol{\Lambda}$. The function $h(x) = x+\frac{1}{x}$, defined for $x>0$, admits 2 as a global minimum for $x=1$, showing that $1-4\alpha+2\alpha(\frac{\lambda_i}{\lambda_j}+\frac{\lambda_j}{\lambda_i}) \geq 1 > 0$. This completes the proof. ∎

The Levi-Civita connection on $\mathcal{M}_{p,k}$ associated with the metric induced by (12), which is crucial when it comes to defining geodesics and Riemannian Hessians, is given in proposition 2.

**Proposition 2.** *Let $\theta = \pi(\bar\theta)\in\mathcal{M}_{p,k}$, $\xi = \mathrm{D}\,\pi(\bar\theta)[\bar\xi]\in T_\theta\mathcal{M}_{p,k}$ and the vector field $\eta = \mathrm{D}\,\pi(\bar\theta)[\bar\eta]$ evaluated at $\theta$, where $\bar\xi$, $\bar\eta\in\mathcal{H}_{\bar\theta}$. The representative $\overline{\nabla_\xi\,\eta}$ in $\mathcal{H}_{\bar\theta}$ of the Levi-Civita connection $\nabla_\xi\,\eta$ on $\mathcal{M}_{p,k}$ is*

$$\overline{\nabla_\xi\,\eta} = P^{\mathcal{H}}_{\bar\theta}(\overline{\nabla}_{\bar\xi}\,\bar\eta),$$

*where $\overline{\nabla}_{\bar\xi}\,\bar\eta$ is the Levi-Civita connection on $\overline{\mathcal{M}}_{p,k}$, given by*

$$\overline{\nabla}_{\bar\xi}\,\bar\eta = P_{\bar\theta}(\mathrm{D}\,\bar\eta[\bar\xi]) + ((\boldsymbol{I}_p - \boldsymbol{UU}^H)\,\mathrm{herm}(\boldsymbol{\eta_U}\boldsymbol{\xi_U}^H)\boldsymbol{U}, -\,\mathrm{herm}(\boldsymbol{\eta_\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}))$$

*Proof:* Let $\bar g_{\bar\theta}(\bar\xi,\bar\eta) = \langle\bar\xi,\bar\eta\rangle_{\bar\theta}$. The Koszul formula [33], which characterizes the Levi-Civita connection, is in our case

$$2\bar g_{\bar\theta}(\overline{\nabla}_{\bar\xi}\,\bar\eta,\bar\nu) - 2\bar g_{\bar\theta}(\mathrm{D}\,\bar\eta[\bar\xi],\bar\nu) = +\,\mathrm{D}\,\bar g_{\bar\theta}[\bar\xi](\bar\eta,\bar\nu) + \mathrm{D}\,\bar g_{\bar\theta}[\bar\eta](\bar\xi,\bar\nu) - \mathrm{D}\,\bar g_{\bar\theta}[\bar\nu](\bar\xi,\bar\eta).$$

To obtain the three terms on the right side of this equation, we have to derive the metric $\bar g_{\bar\theta}$ with respect to $\bar\theta$. One can check that

$$\begin{aligned}\mathrm{D}\,\bar g_{\bar\theta}[\bar\nu](\bar\xi,\bar\eta) = &-\Re\mathfrak{e}(\mathrm{tr}(\boldsymbol{\xi_U}^H\,\mathrm{herm}(\boldsymbol{U}\boldsymbol{\nu_U}^H)\boldsymbol{\eta_U}))\\ &-\beta\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu_\Sigma})\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta_\Sigma})\\ &-\beta\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma})\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta_\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu_\Sigma})\\ &-2\alpha\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta_\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu_\Sigma}).\end{aligned}$$

It follows that the right side of the Koszul formula is

$$\begin{aligned}&\mathrm{D}\,\bar g_{\bar\theta}[\bar\xi](\bar\eta,\bar\nu) + \mathrm{D}\,\bar g_{\bar\theta}[\bar\eta](\bar\xi,\bar\nu) - \mathrm{D}\,\bar g_{\bar\theta}[\bar\nu](\bar\xi,\bar\eta) =\\ &\quad\mathrm{tr}(\boldsymbol{\nu_U}^H(2\,\mathrm{herm}(\boldsymbol{\eta_U}\boldsymbol{\xi_U}^H)\boldsymbol{U} - \boldsymbol{U}\,\mathrm{herm}(\boldsymbol{\eta_U}^H\boldsymbol{\xi_U})))\\ &\quad-2\alpha\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta_\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu_\Sigma})\\ &\quad-2\beta\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta_\Sigma})\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu_\Sigma}).\end{aligned}$$

Moreover,

$$\mathrm{tr}(\boldsymbol{\nu_U}^H\widetilde{\boldsymbol{Z}}_{\boldsymbol{U}}) = \mathrm{tr}(\boldsymbol{\nu_U}^H(\boldsymbol{I}_p - \frac{1}{2}\boldsymbol{UU}^H)(\boldsymbol{I}_p + \boldsymbol{UU}^H)\widetilde{\boldsymbol{Z}}_{\boldsymbol{U}}).$$

It follows that

$$\mathrm{D}\,\bar g_{\bar\theta}[\bar\xi](\bar\eta,\bar\nu) + \mathrm{D}\,\bar g_{\bar\theta}[\bar\eta](\bar\xi,\bar\nu) - \mathrm{D}\,\bar g_{\bar\theta}[\bar\nu](\bar\xi,\bar\eta) = 2\bar g_{\bar\theta}(\overline{Z},\bar\nu),$$

where

$$\overline{Z} = ((\boldsymbol{I}_p + \boldsymbol{UU}^H)\,\mathrm{herm}(\boldsymbol{\eta_U}\boldsymbol{\xi_U}^H)\boldsymbol{U} - \frac{1}{2}\boldsymbol{U}\,\mathrm{herm}(\boldsymbol{\eta_U}^H\boldsymbol{\xi_U}), -\,\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta_\Sigma}).$$

Since $\bar{\nu} \in T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k}$ and the projection map (13) is orthogonal according to (12), projecting $\overline{Z}$ on $T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k}$ does not change the metric, *i.e.*, $\bar{g}_{\bar{\theta}}(\overline{Z},\bar{\nu}) = \bar{g}_{\bar{\theta}}(P_{\bar{\theta}}(\overline{Z}),\bar{\nu})$. Thus,

$$\mathrm{D}\,\bar{g}_{\bar{\theta}}[\bar{\xi}](\bar{\eta},\bar{\nu}) + \mathrm{D}\,\bar{g}_{\bar{\theta}}[\bar{\eta}](\bar{\xi},\bar{\nu}) - \mathrm{D}\,\bar{g}_{\bar{\theta}}[\bar{\nu}](\bar{\xi},\bar{\eta}) = 2\bar{g}_{\bar{\theta}}((\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}^H)\,\mathrm{herm}(\boldsymbol{\eta_U}\boldsymbol{\xi}_{\boldsymbol{U}}^H)\boldsymbol{U}, -\,\mathrm{herm}(\boldsymbol{\eta_\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}),\bar{\nu}).$$

The same way, $\bar{g}_{\bar{\theta}}(\mathrm{D}\,\bar{\eta}[\bar{\xi}],\bar{\nu}) = \bar{g}_{\bar{\theta}}(P_{\bar{\theta}}(\mathrm{D}\,\bar{\eta}[\bar{\xi}]),\bar{\nu})$. Injecting these results in the Koszul formula, the Levi-Civita connection $\overline{\nabla}_{\bar{\xi}}\,\bar{\eta}$ on $\overline{\mathcal{M}}_{p,k}$ is finally obtained by identification. The Levi-Civita connection $\nabla_{\xi}\,\eta$ on $\mathcal{M}_{p,k}$ is then simply given by [33, proposition 5.3.3]. ■

The geodesics in $\mathcal{M}_{p,k}$ associated with the metric induced by (12) are given in proposition 3. Unfortunately, an analytical formula for the geodesic between two points $\theta$ and $\widehat{\theta}$ in $\mathcal{M}_{p,k}$ is not known. As a direct consequence, the Riemannian logarithm map and the Riemannian distance function on $\mathcal{M}_{p,k}$ are not known in closed form.

**Proposition 3.** *Let $\theta = \pi(\bar{\theta}) \in \mathcal{M}_{p,k}$ and $\xi = \mathrm{D}\,\pi(\bar{\theta})[\bar{\xi}] \in T_{\theta}\mathcal{M}_{p,k}$, where $\bar{\xi} \in \mathcal{H}_{\bar{\theta}}$. The representative in $\overline{\mathcal{M}}_{p,k}$ of the geodesic in $\mathcal{M}_{p,k}$ associated with the metric induced by* (12) *starting at $\theta$ in the direction $\xi$ is*[6]

$$\bar{\gamma}(t) = (\boldsymbol{U}(t),\boldsymbol{\Sigma}(t)) = \left([\boldsymbol{U}\,\boldsymbol{Q}]\exp t\begin{pmatrix}\boldsymbol{U}^H\boldsymbol{\xi_U} & -\boldsymbol{R}^H\\ \boldsymbol{R} & \boldsymbol{0}\end{pmatrix}\begin{bmatrix}\boldsymbol{I}_k\\ \boldsymbol{0}\end{bmatrix},\ \boldsymbol{\Sigma}^{1/2}\exp(t\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1/2})\boldsymbol{\Sigma}^{1/2}\right),$$

*where $\boldsymbol{Q}$ and $\boldsymbol{R}$ correspond to the QR decomposition of $(\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}^H)\boldsymbol{\xi_U}$.*

*Proof:* A direct proof that $\bar{\gamma}(t)$ is a geodesic in $\overline{\mathcal{M}}_{p,k}$ consists in verifying that it is solution of the differential equation $\overline{\nabla}_{\dot{\bar{\gamma}}(t)}\,\dot{\bar{\gamma}}(t) = \boldsymbol{0}$, where $\dot{\bar{\gamma}}(t)$ is the derivative of $\bar{\gamma}(t)$. However, it is enough to argue that $\boldsymbol{U}(t)$ corresponds to the geodesic in $\mathrm{St}_{p,k}$ equiped with its canonical metric [28] and $\boldsymbol{\Sigma}(t)$ is the geodesic in $\mathcal{H}_k^{++}$ equiped with the considered affine invariant metric; see *e.g.*, [29], [47].

To show that $\bar{\gamma}(t)$ is a proper representative of the geodesic in $\mathcal{M}_{p,k}$, as $\pi$ is a Riemannian submersion, it suffices to show that $\bar{\gamma}(t)$ stays horizontal in $\overline{\mathcal{M}}_{p,k}$, *i.e.*, $\dot{\bar{\gamma}}(t) \in \mathcal{H}_{\bar{\gamma}(t)}$ [42, proposition 2.109]. One can check that $\boldsymbol{U}(t)^H\dot{\boldsymbol{U}}(t) = \boldsymbol{U}^H\boldsymbol{\xi_U}$, $\boldsymbol{\Sigma}(t)^{-1}\dot{\boldsymbol{\Sigma}}(t) = \boldsymbol{\Sigma}^{-1}\boldsymbol{\xi_\Sigma}$ and $\dot{\boldsymbol{\Sigma}}(t)\boldsymbol{\Sigma}(t)^{-1} = \boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1}$, which is enough to conclude. ■

## V. Riemannian optimization for robust covariance estimation

We build a Riemannian optimization framework on $\mathcal{M}_{p,k}$ for robust estimation of covariance matrices admitting the structure (4). In section V-A, we provide the objects required to perform Riemannian optimization (see section II) on $\mathcal{M}_{p,k}$, *i.e.*, the Riemannian gradient and Hessian and a retraction, which corresponds to a second-order approximation of the geodesics of proposition 3. In section V-B, we develop tools to treat the family of cost functions of interest, which are originally defined on $\mathcal{H}_p^{++}$. In particular, we deal with Tyler's $M$-estimator cost function defined in (10).

[6] The considered geodesic $\boldsymbol{U}(t)$ on $\mathrm{St}_{p,k}$ is optimal (from a dimensionality point of view) only if $k \leq p/2$. If $k > p/2$, it is more advantageous to replace $\boldsymbol{Q}$ with $\boldsymbol{U}_\perp$ and $\boldsymbol{R}$ with $\boldsymbol{U}_\perp^H\boldsymbol{\xi_U}$, where $\boldsymbol{U}_\perp \in \mathrm{St}_{p,p-k}$ such that $\boldsymbol{U}^H\boldsymbol{U}_\perp = \boldsymbol{0}$; see [28].

### A. Riemannian optimization on $\mathcal{M}_{p,k}$

Let $\bar{f} : \overline{\mathcal{M}}_{p,k} \to \mathbb{R}$ be an objective function that induces a function $f$ on the quotient $\mathcal{M}_{p,k}$, *i.e.*, $\bar{f}$ is invariant along the equivalence classes (8): for all $\bar{\theta} \in \overline{\mathcal{M}}_{p,k}$ and $\boldsymbol{O} \in \mathcal{U}_k$, $\bar{f}(\bar{\theta}) = \bar{f}(\phi_{\boldsymbol{O}}(\bar{\theta}))$, where $\phi_{\boldsymbol{O}}(\bar{\theta}) = (\boldsymbol{U}\boldsymbol{O},\boldsymbol{O}^H\boldsymbol{\Sigma}\boldsymbol{O})$, as in section IV. To perform Riemannian optimization, it remains to define the Riemannian gradient and Hessian of $f$ along with a retraction on $\mathcal{M}_{p,k}$. Proposition 4 provides formulas to compute the Riemannian gradient and Hessian of $f$ on $\mathcal{M}_{p,k}$ from the Euclidean gradient and Hessian of $\bar{f}$ on $\overline{\mathcal{M}}_{p,k}$.

**Proposition 4.** *Given $\theta = \pi(\bar{\theta}) \in \mathcal{M}_{p,k}$, the representative in $\mathcal{H}_{\bar{\theta}}$ of the Riemannian gradient of $f$ at $\theta$ is the Riemannian gradient of $\bar{f}$ at $\bar{\theta}$, which is*

$$\mathrm{grad}_{\overline{\mathcal{M}}_{p,k}}\,\bar{f}(\bar{\theta}) = \left(\boldsymbol{G_U} - \boldsymbol{U}\boldsymbol{G}_{\boldsymbol{U}}^H\boldsymbol{U},\ \frac{\boldsymbol{\Sigma}\,\mathrm{herm}(\boldsymbol{G_\Sigma})\boldsymbol{\Sigma}}{\alpha} - \frac{\beta\,\mathrm{tr}(\boldsymbol{G_\Sigma}\boldsymbol{\Sigma})}{\alpha(\alpha+k\beta)}\boldsymbol{\Sigma}\right),$$

*where $\mathrm{grad}_{\mathcal{E}}\,\bar{f}(\bar{\theta}) = (\boldsymbol{G_U},\boldsymbol{G_\Sigma})$ is the Euclidean gradient of $\bar{f}$ in $\mathbb{C}^{p\times k}\times\mathbb{C}^{k\times k}$.*

*Given $\xi = \mathrm{D}\,\pi(\bar{\theta})[\bar{\xi}] \in T_{\theta}\mathcal{M}_{p,k}$, the representative in $\mathcal{H}_{\bar{\theta}}$ of the Riemannian Hessian $\mathrm{Hess}_{\mathcal{M}_{p,k}}\,f(\theta)[\xi]$ of $f$ at $\theta$ in direction $\xi$ is*

$$\overline{\mathrm{Hess}_{\mathcal{M}_{p,k}}\,f(\theta)[\xi]} = P_{\bar{\theta}}^{\mathcal{H}}(\mathrm{Hess}_{\overline{\mathcal{M}}_{p,k}}\,\bar{f}(\bar{\theta})[\bar{\xi}]),$$

*where $\mathrm{Hess}_{\overline{\mathcal{M}}_{p,k}}\,\bar{f}(\bar{\theta})[\bar{\xi}]$ is the Riemannian Hessian of $\bar{f}$ at $\bar{\theta}$ in direction $\bar{\xi}$, given by*

$$\begin{aligned}\mathrm{Hess}_{\overline{\mathcal{M}}_{p,k}}\,\bar{f}(\bar{\theta})[\bar{\xi}] = \Big(&\boldsymbol{H_U} - \boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}^H\boldsymbol{U} - \boldsymbol{U}\,\mathrm{skew}(\boldsymbol{G}_{\boldsymbol{U}}^H\boldsymbol{\xi_U}) \\ &- \mathrm{skew}(\boldsymbol{G_U}\boldsymbol{\xi}_{\boldsymbol{U}}^H)\boldsymbol{U} - \frac{1}{2}(\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}^H)\boldsymbol{\xi_U}\boldsymbol{U}^H\boldsymbol{G_U}, \\ &\frac{1}{\alpha}(\boldsymbol{\Sigma}\,\mathrm{herm}(\boldsymbol{H_\Sigma})\boldsymbol{\Sigma} + \mathrm{herm}(\boldsymbol{\Sigma}\,\mathrm{herm}(\boldsymbol{G_\Sigma})\boldsymbol{\xi_\Sigma})) \\ &- \frac{\beta\,\mathrm{tr}(\boldsymbol{H_\Sigma}\boldsymbol{\Sigma} + \boldsymbol{G_\Sigma}\boldsymbol{\xi_\Sigma})}{\alpha(\alpha+k\beta)}\boldsymbol{\Sigma}\Big),\end{aligned}$$

*where* skew *returns the skew-Hermitian part of its argument and $\mathrm{Hess}_{\mathcal{E}}\,\bar{f}(\bar{\theta})[\bar{\xi}] = (\boldsymbol{H_U},\boldsymbol{H_\Sigma})$ is the Euclidean Hessian of $\bar{f}$ at $\bar{\theta}$ in direction $\bar{\xi}$, i.e., $\mathrm{Hess}_{\mathcal{E}}\,\bar{f}(\bar{\theta})[\bar{\xi}] = \mathrm{D}\,\mathrm{grad}_{\mathcal{E}}\,\bar{f}(\bar{\theta})[\bar{\xi}]$.*

*Proof:* The Riemannian and Euclidean gradients of $\bar{f}$ at $\bar{\theta}$ are defined by

$$\mathrm{D}\,\bar{f}(\bar{\theta})[\bar{\xi}] = \langle\mathrm{grad}_{\overline{\mathcal{M}}_{p,k}}\,\bar{f}(\bar{\theta}),\bar{\xi}\rangle_{\bar{\theta}} = \langle\mathrm{grad}_{\mathcal{E}}\,\bar{f}(\bar{\theta}),\bar{\xi}\rangle^{\mathcal{E}},$$

where $\langle\cdot,\cdot\rangle^{\mathcal{E}}$ is the Euclidean metric on $\mathbb{C}^{p\times k}\times\mathbb{C}^{k\times k}$, which is given by

$$\langle\xi,\eta\rangle^{\mathcal{E}} = \mathfrak{Re}(\mathrm{tr}(\boldsymbol{\xi}_{\boldsymbol{U}}^H\boldsymbol{\eta_U}) + \mathrm{tr}(\boldsymbol{\xi}_{\boldsymbol{\Sigma}}^H\boldsymbol{\eta_\Sigma})). \tag{14}$$

Injecting the proposed formula for the gradient $\mathrm{grad}_{\overline{\mathcal{M}}_{p,k}}\,\bar{f}(\bar{\theta})$ in the metric (12) shows that $\langle\mathrm{grad}_{\overline{\mathcal{M}}_{p,k}}\,\bar{f}(\bar{\theta}),\bar{\xi}\rangle_{\bar{\theta}}$ is equal to $\langle\mathrm{grad}_{\mathcal{E}}\,\bar{f}(\bar{\theta}),\bar{\xi}\rangle^{\mathcal{E}}$. To show that it is the Riemannian gradient of $\bar{f}$ at $\bar{\theta} \in \overline{\mathcal{M}}_{p,k}$, we also need to check that it belongs to $T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k}$ defined in (11). It is readily checked that the component corresponding to $\boldsymbol{\Sigma}$ is Hermitian and that

$$\boldsymbol{U}^H(\boldsymbol{G_U} - \boldsymbol{U}\boldsymbol{G_U^H}\boldsymbol{U}) + (\boldsymbol{G_U} - \boldsymbol{U}\boldsymbol{G_U^H}\boldsymbol{U})^H\boldsymbol{U} = \boldsymbol{0},$$

which is enough to conclude. From section II, we further know that it belongs to $\mathcal{H}_{\bar{\theta}}$ and is the representative of the Riemannian gradient of $f$ at $\theta \in \mathcal{M}_{p,k}$.

Recall that the Riemannian Hessian of $\bar{f}$ at $\bar{\theta}$ in direction $\bar{\xi}$ is defined as $\mathrm{Hess}_{\overline{\mathcal{M}}_{p,k}} \bar{f}(\bar{\theta})[\bar{\xi}] = \overline{\nabla}_{\bar{\xi}} \, \mathrm{grad}_{\overline{\mathcal{M}}_{p,k}} \bar{f}(\bar{\theta})$. The result is obtained by plugging the formula of the gradient in the one of the Levi-Civita connection $\overline{\nabla}$ on $\overline{\mathcal{M}}_{p,k}$ defined in proposition 2. Finally, the representative of the Riemannian Hessian of $f$ at $\theta = \pi(\bar{\theta})$ in direction $\xi = \mathrm{D}\,\pi(\bar{\theta})[\bar{\xi}]$ is obtained by definition of the Levi-Civita connection $\nabla$ on $\mathcal{M}_{p,k}$, given in proposition 2. ■

It only remains to provide a retraction on $\mathcal{M}_{p,k}$. As explained in section II, a solution is to take the Riemannian exponential map defined through the geodesics of proposition 3. However, for numerical stability reasons (the exponential function goes quickly towards infinity), we rather choose a second order approximation of this exponential map, which, for $\theta = \pi(\bar{\theta}) \in \mathcal{M}_{p,k}$ and $\xi = \mathrm{D}\,\pi(\bar{\theta})[\bar{\xi}] \in T_\theta\mathcal{M}_{p,k}$, is represented by

$$\bar{R}_{\bar{\theta}}(\bar{\xi}) = \left( [\boldsymbol{U}\,\boldsymbol{Q}] \,\mathrm{uf} \circ \Gamma \begin{pmatrix} \boldsymbol{U}^H\boldsymbol{\xi_U} & -\boldsymbol{R}^H \\ \boldsymbol{R} & \boldsymbol{0} \end{pmatrix} \begin{bmatrix} \boldsymbol{I}_k \\ \boldsymbol{0} \end{bmatrix} \right.$$
$$\left. \boldsymbol{\Sigma}^{1/2}\Gamma(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\xi_\Sigma}\boldsymbol{\Sigma}^{-1/2})\boldsymbol{\Sigma}^{1/2} \right), \quad (15)$$

where uf returns the orthogonal factor of the polar decomposition and $\Gamma(\boldsymbol{X}) = \boldsymbol{I} + \boldsymbol{X} + \frac{1}{2}\boldsymbol{X}^2$ is a second order approximation of the matrix exponential.

### B. Robust covariance estimation

In this section, we propose new estimation procedures that leverage the Riemannian optimization framework of the previous section. Recall that we aim at estimating covariance matrices admitting the structure $\boldsymbol{R} = \boldsymbol{I}_p + \varphi(\theta)$, where $\varphi(\theta) = \bar{\varphi}(\bar{\theta})$ as in (6)), from $n$ samples $\{\boldsymbol{x}_i\}$ drawn from a complex elliptically symmetric distribution (cf. section III). To that end, we are interested in objective functions $L : \mathcal{M}_{p,k} \to \mathbb{R}$ which have the form

$$L(\theta) = L^{++}(\boldsymbol{I}_p + \varphi(\theta)), \quad (16)$$

where $L^{++} : \mathcal{H}_p^{++} \to \mathbb{R}$ corresponds to a likelihood function as in (2). To perform Riemannian optimization of $L$ with the tools developed in section V-A, we simply need to have the Euclidean gradient and Hessian of $\bar{L} = L \circ \pi$. Proposition 5 shows that they can be obtained from those of $L^{++}$. For the Hessian, we need the directional derivative of $\bar{\varphi}$ at $\bar{\theta}$, which is given, for all $\bar{\xi} \in T_{\bar{\theta}}\overline{\mathcal{M}}_{p,k}$, by

$$\mathrm{D}\,\bar{\varphi}(\bar{\theta})[\bar{\xi}] = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\xi}_{\boldsymbol{U}}^H + \boldsymbol{\xi_U}\boldsymbol{\Sigma}\boldsymbol{U}^H + \boldsymbol{U}\boldsymbol{\xi_\Sigma}\boldsymbol{U}^H. \quad (17)$$

**Proposition 5.** *The Euclidean gradient of $\bar{L} = L \circ \pi$ at $\bar{\theta} \in \overline{\mathcal{M}}_{p,k}$ is given by*

$$\mathrm{grad}_{\mathcal{E}}\, \bar{L}(\bar{\theta}) = (2\boldsymbol{G}_{\bar{\theta}}^{++}\boldsymbol{U}\boldsymbol{\Sigma}, \boldsymbol{U}^H\boldsymbol{G}_{\bar{\theta}}^{++}\boldsymbol{U}),$$

*where $\boldsymbol{G}_{\bar{\theta}}^{++} = \mathrm{grad}_{\mathfrak{E}}\, L^{++}(\boldsymbol{I}_p + \bar{\varphi}(\bar{\theta}))$ is the Euclidean gradient of $L^{++}$ at $\boldsymbol{I}_p + \bar{\varphi}(\bar{\theta}) \in \mathcal{H}_p^{++}$, with $\bar{\varphi}(\bar{\theta})$ defined in* (6).

*The Euclidean Hessian of $\bar{L}$ at $\bar{\theta}$ in direction $\bar{\xi}$ is*

$$\mathrm{Hess}_{\mathcal{E}}\, \bar{L}(\bar{\theta})[\bar{\xi}] = (2\boldsymbol{H}_{\bar{\theta}}^{++}\boldsymbol{U}\boldsymbol{\Sigma} + 2\boldsymbol{G}_{\bar{\theta}}^{++}(\boldsymbol{\xi_U}\boldsymbol{\Sigma} + \boldsymbol{U}\boldsymbol{\xi_\Sigma}),$$
$$\boldsymbol{U}^H\boldsymbol{H}_{\bar{\theta}}^{++}\boldsymbol{U} + \boldsymbol{U}^H\boldsymbol{G}_{\bar{\theta}}^{++}\boldsymbol{\xi_U} + \boldsymbol{\xi}_{\boldsymbol{U}}^H\boldsymbol{G}_{\bar{\theta}}^{++}\boldsymbol{U}),$$

*where $\boldsymbol{H}_{\bar{\theta}}^{++} = \mathrm{Hess}_{\mathfrak{E}}\, L^{++}(\boldsymbol{I}_p + \bar{\varphi}(\bar{\theta}))[\mathrm{D}\,\bar{\varphi}(\bar{\theta})[\bar{\xi}]]$ is the Euclidean Hessian of $L^{++}$ at $\boldsymbol{I}_p + \bar{\varphi}(\bar{\theta}) \in \mathcal{H}_p^{++}$ in direction $\mathrm{D}\,\bar{\varphi}(\bar{\theta})[\bar{\xi}] \in \mathcal{H}_p$, which is defined in* (17).

*Proof:* Let $\mathrm{grad}_{\mathcal{E}}\, \bar{L}(\bar{\theta}) = (\boldsymbol{G_U}, \boldsymbol{G_\Sigma})$. By definition,

$$\mathrm{D}\,\bar{f}(\bar{\theta})[\bar{\xi}] = \langle \mathrm{grad}_{\mathcal{E}}\, \bar{L}(\bar{\theta}), \bar{\xi} \rangle^{\mathcal{E}},$$

where $\langle \cdot, \cdot \rangle^{\mathcal{E}}$ is defined in (14). We also have

$$\begin{aligned} \mathrm{D}\,\bar{f}(\bar{\theta})[\bar{\xi}] &= \mathrm{D}\,f^{++}(\boldsymbol{I}_p + \bar{\varphi}(\bar{\theta}))[\mathrm{D}\,\bar{\varphi}(\bar{\theta})[\bar{\xi}]] \\ &= \langle \boldsymbol{G}_{\bar{\theta}}^{++}, \mathrm{D}\,\bar{\varphi}(\bar{\theta})[\bar{\xi}] \rangle^{\mathfrak{E}}, \end{aligned}$$

where $\langle \cdot, \cdot \rangle^{\mathfrak{E}}$ is the Euclidean metric on $\mathbb{C}^{p \times p}$, which is

$$\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle^{\mathfrak{E}} = \mathfrak{Re}(\mathrm{tr}(\boldsymbol{\xi}^H\boldsymbol{\eta})).$$

We thus need to show that

$$\mathfrak{Re}(\mathrm{tr}(\boldsymbol{G}_{\boldsymbol{U}}^H\boldsymbol{\xi_U}) + \mathrm{tr}(\boldsymbol{G}_{\boldsymbol{\Sigma}}^H\boldsymbol{\xi_\Sigma})) = \mathfrak{Re}(\mathrm{tr}(\boldsymbol{G}_{\bar{\theta}}^{++\,H}\,\mathrm{D}\,\bar{\varphi}(\bar{\theta})[\bar{\xi}])).$$

It is achieved by plugging the proposed formula for the Euclidean gradient $\mathrm{grad}_{\mathcal{E}}\, \bar{L}(\bar{\theta}) = (\boldsymbol{G_U}, \boldsymbol{G_\Sigma})$ and the definition of $\mathrm{D}\,\bar{\varphi}(\bar{\theta})[\bar{\xi}]$ provided in (17). The Hessian is defined as $\mathrm{Hess}_{\mathcal{E}}\, \bar{L}(\bar{\theta})[\bar{\xi}] = \mathrm{D}\,\mathrm{grad}_{\mathcal{E}}\, \bar{L}(\bar{\theta})[\bar{\xi}]$. The proposed formula follows from basic calculations. ■

Note that, in practice, the density generator $g$ (or any scaling ambiguity as discussed in remark 1) is unknown. In order to propose a robust estimation method, we focus on Tyler's $M$-estimator cost function in (10) due to its "distribution-free" property [8]. Moreover, this function is scale invariant, i.e., $L_{\mathrm{T}}^{++}(\boldsymbol{R}) = L_{\mathrm{T}}^{++}(c\boldsymbol{R})$, $\forall c \in \mathbb{R}^*$. Thus, from the chosen parameterization's inherent scale, it is therefore noted that the proposed algorithms will only produce estimates of the *shape* of the covariance matrix. To be able to compute Tyler's $M$-estimator on $\mathcal{M}_{p,k}$ from minimizing $L_{\mathrm{T}}$, it remains to give the Euclidean gradient and Hessian of $L_{\mathrm{T}}^{++}$ defined in (3). To do so, we define $\Psi : \mathcal{H}_p^{++} \to \mathcal{H}_p$ and its directional derivative as

$$\begin{aligned} \Psi(\boldsymbol{R}) &= \textstyle\sum_i \frac{\boldsymbol{x}_i\boldsymbol{x}_i^H}{\boldsymbol{x}_i^H\boldsymbol{R}^{-1}\boldsymbol{x}_i}, \\ \mathrm{D}\,\Psi(\boldsymbol{R})[\boldsymbol{\xi_R}] &= \textstyle\sum_i \frac{\boldsymbol{x}_i^H\boldsymbol{R}^{-1}\boldsymbol{\xi_R}\boldsymbol{R}^{-1}\boldsymbol{x}_i}{(\boldsymbol{x}_i^H\boldsymbol{R}^{-1}\boldsymbol{x}_i)^2}\boldsymbol{x}_i\boldsymbol{x}_i^H. \end{aligned}$$

It follows that the Euclidean gradient of $L_{\mathrm{T}}^{++}$ at $\boldsymbol{R} \in \mathcal{H}_p^{++}$ is

$$\mathrm{grad}_{\mathfrak{E}}\, L_{\mathrm{T}}^{++}(\boldsymbol{R}) = \boldsymbol{R}^{-1}(n\boldsymbol{R} - p\Psi(\boldsymbol{R}))\boldsymbol{R}^{-1}, \quad (18)$$

and the Euclidean Hessian of $L_{\mathrm{T}}^{++}$ at $\boldsymbol{R} \in \mathcal{H}_p^{++}$ in direction $\boldsymbol{\xi_R} \in \mathcal{H}_p$ is

$$\mathrm{Hess}_{\mathfrak{E}}\, L_{\mathrm{T}}^{++}(\boldsymbol{R})[\boldsymbol{\xi_R}] = 2p\boldsymbol{R}^{-1}\,\mathrm{herm}(\boldsymbol{\xi_R}\boldsymbol{R}^{-1}\Psi(\boldsymbol{R}))\boldsymbol{R}^{-1}$$
$$- \boldsymbol{R}^{-1}(p\,\mathrm{D}\,\Psi(\boldsymbol{R})[\boldsymbol{\xi_R}] + n\boldsymbol{\xi_R})\boldsymbol{R}^{-1}. \quad (19)$$

## VI. Intrinsic Cramér-Rao lower bounds for low-rank structured elliptical models

Intrinsic Cramér-Rao bounds on $\mathcal{M}_{p,k}$ for the estimation problem of low-rank structured complex elliptically symmetric distributions are studied. First, in section VI-A, the Fisher information metric of the complex elliptically symmetric distribution on $\mathcal{M}_{p,k}$ is derived. In section VI-B, the usual intrinsic Cramér-Rao bound framework associated with the Riemannian geometry of section IV on $\mathcal{M}_{p,k}$ is investigated. Since the corresponding estimation error is hard to evaluate in practice, section VI-C, addresses the issue by developing a new alternative intrinsic Cramér-Rao bound beyond the Riemannian geometry framework (i.e. on the proposed divergence). Finally, in section VI-D, an intrinsic Cramér-Rao bound on the Grassmann manifold $\mathcal{G}_{p,k}$ for subspace recovery is derived.

### A. Fisher information metric

Before deriving the Fisher information metric on $\mathcal{M}_{p,k}$ of complex elliptically symmetric distributions with covariance matrix of the form (4), we first give the general proposition 6. The Fisher information metric on $\overline{\mathcal{M}}_{p,k}$ (which induces the one on $\mathcal{M}_{p,k}$) of probability density function (9) is then obtained from the one of probability density function (1) on $\mathcal{H}_p^{++}$ in corollary 1.

**Proposition 6.** *Let two manifolds $\mathcal{M}$, $\mathcal{N}$ and the smooth mapping $\psi : \mathcal{M} \to \mathcal{N}$. Consider the probability density function on $\mathcal{M}$*

$$f^{\mathcal{M}}(x|\theta) = f^{\mathcal{N}}(x|\psi(\theta)),$$

*where $f^{\mathcal{N}}$ is a probability density function on $\mathcal{N}$ whose Fisher information metric is $g^{\mathcal{N}}$. It follows that the Fisher information metric $g^{\mathcal{M}}$ on $\mathcal{M}$ associated with $f^{\mathcal{M}}$ is, given $\theta \in \mathcal{M}$ and $\xi$, $\eta \in T_\theta\mathcal{M}$,*

$$g_\theta^{\mathcal{M}}(\xi,\eta) = g_{\psi(\theta)}^{\mathcal{N}}(\mathrm{D}\,\psi(\theta)[\xi], \mathrm{D}\,\psi(\theta)[\eta]).$$

*Proof:* Let $L_x^{\mathcal{M}}(\theta)$ the log-likelihood on $\mathcal{M}$ of $f^{\mathcal{M}}(x|\theta)$. By definition, $L_x^{\mathcal{M}}(\theta) = L_x^{\mathcal{N}}(\psi(\theta))$ and

$$\begin{aligned} g_\theta^{\mathcal{M}}(\xi,\eta) &= \mathbb{E}_x\left[\mathrm{D}\,L_x^{\mathcal{M}}(\theta)[\xi]\,\mathrm{D}\,L_x^{\mathcal{M}}(\theta)[\eta]\right] \\ &= \mathbb{E}_x\left[\mathrm{D}\,L_x^{\mathcal{N}}(\psi(\theta))[\mathrm{D}\,\psi(\theta)[\xi]]\,\mathrm{D}\,L_x^{\mathcal{N}}(\psi(\theta))[\mathrm{D}\,\psi(\theta)[\eta]]\right] \\ &= g_{\psi(\theta)}^{\mathcal{N}}(\mathrm{D}\,\psi(\theta)[\xi], \mathrm{D}\,\psi(\theta)[\eta]). \end{aligned}$$

■

**Corollary 1.** *The Fisher information metric on $\overline{\mathcal{M}}_{p,k}$ corresponding to the probability density function (9) is, for $\bar\theta \in \overline{\mathcal{M}}_{p,k}$ and $\bar\xi$, $\bar\eta \in T_{\bar\theta}\overline{\mathcal{M}}_{p,k}$,*

$$g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi,\bar\eta) = g_{\boldsymbol{I}_p + \overline{\varphi}(\bar\theta)}^{\mathcal{H}_p^{++}}(\mathrm{D}\,\overline{\varphi}(\bar\theta)[\bar\xi], \mathrm{D}\,\overline{\varphi}(\bar\theta)[\bar\eta]),$$

*where $\overline{\varphi}(\bar\theta)$ and $\mathrm{D}\,\overline{\varphi}(\bar\theta)[\bar\xi]$ are defined in (6) and (17), and*

$$\begin{aligned} g_{\boldsymbol{R}}^{\mathcal{H}_p^{++}}(\boldsymbol{\xi}_{\boldsymbol{R}},\boldsymbol{\eta}_{\boldsymbol{R}}) &= n\alpha^{++}\operatorname{tr}(\boldsymbol{R}^{-1}\boldsymbol{\xi}_{\boldsymbol{R}}\boldsymbol{R}^{-1}\boldsymbol{\eta}_{\boldsymbol{R}}) \\ &\quad + n(\alpha^{++}-1)\operatorname{tr}(\boldsymbol{R}^{-1}\boldsymbol{\xi}_{\boldsymbol{R}})\operatorname{tr}(\boldsymbol{R}^{-1}\boldsymbol{\xi}_{\boldsymbol{R}}) \end{aligned}$$

*is the Fisher information on $\mathcal{H}_p^{++}$ associated with the probability density function (1), where $\alpha^{++}$ is a scalar that only depends on the density generator $g$ in (1) [29].*

### B. Intrinsic Cramér-Rao bound associated to the Riemannian geometry on $\mathcal{M}_{p,k}$

In order to employ the intrinsic Cramér-Rao bound framework presented in section II, an orthonormal basis of $T_{\bar\theta}\overline{\mathcal{M}}_{p,k}$ is required. Such basis is given in proposition 7.

**Proposition 7.** *Given $\bar\theta \in \overline{\mathcal{M}}_{p,k}$, an orthonormal basis $\{e_q\}_{1\le q\le 2pk}$ of the tangent space $T_{\bar\theta}\overline{\mathcal{M}}_{p,k}$ is given by*

$$\Big\{\{(\boldsymbol{e}_{\boldsymbol{U}_\perp}^{ij},\boldsymbol{0}),(\widetilde{\boldsymbol{e}}_{\boldsymbol{U}_\perp}^{ij},\boldsymbol{0})\}_{\substack{1\le i\le p-k \\ 1\le j\le k}}, \{(\boldsymbol{e}_{\boldsymbol{U}}^{ij},\boldsymbol{0})\}_{1\le j<i\le k},$$
$$\{(\widetilde{\boldsymbol{e}}_{\boldsymbol{U}}^{ij},\boldsymbol{0})\}_{1\le j\le i\le k}, \{(\boldsymbol{0},\boldsymbol{e}_{\boldsymbol{\Sigma}}^{ij})\}_{1\le j\le i\le k}, \{(\boldsymbol{0},\widetilde{\boldsymbol{e}}_{\boldsymbol{\Sigma}}^{ij})\}_{1\le j<i\le k}\Big\},$$

*where*

- $\boldsymbol{e}_{\boldsymbol{U}_\perp}^{ij} = \boldsymbol{U}_\perp\boldsymbol{K}^{ij}$, $\widetilde{\boldsymbol{e}}_{\boldsymbol{U}_\perp}^{ij} = \mathrm{i}\boldsymbol{U}_\perp\boldsymbol{K}^{ij}$*: $\boldsymbol{U}_\perp \in \mathrm{St}_{p,p-k}$, $\boldsymbol{U}^H\boldsymbol{U}_\perp = \boldsymbol{0}$; $\boldsymbol{K}^{ij} \in \mathbb{R}^{(p-k)\times k}$, its $ij^{\text{th}}$ element is 1, zeros elsewhere.*
- $\boldsymbol{e}_{\boldsymbol{U}}^{ij} = \boldsymbol{U}\boldsymbol{\Omega}^{ij}$*: $\boldsymbol{\Omega}^{ij} \in \mathcal{H}_k^\perp$, its $ij^{\text{th}}$ and $ji^{\text{th}}$ elements are 1 and $-1$, zeros elsewhere.*
- $\widetilde{\boldsymbol{e}}_{\boldsymbol{U}}^{ij} = \boldsymbol{U}\widetilde{\boldsymbol{\Omega}}^{ij}$*: $\widetilde{\boldsymbol{\Omega}}^{ii} \in \mathcal{H}_k^\perp$, its $ii^{\text{th}}$ element is $\sqrt{2}\mathrm{i}$, zeros elsewhere; $\widetilde{\boldsymbol{\Omega}}^{ij} \in \mathcal{H}_k^\perp$, $i > j$, its $ij^{\text{th}}$ and $ji^{\text{th}}$ elements are $\mathrm{i}$, zeros elsewhere.*
- $\boldsymbol{e}_{\boldsymbol{\Sigma}}^{ij} = \frac{1}{\sqrt{\alpha}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{H}^{ij}\boldsymbol{\Sigma}^{1/2} + \frac{\sqrt{\alpha}-\sqrt{\alpha+k\beta}}{k\sqrt{\alpha}\sqrt{\alpha+k\beta}}\operatorname{tr}(\boldsymbol{H}^{ij})\boldsymbol{\Sigma}$*: $\boldsymbol{H}^{ii} \in \mathcal{H}_k$, its $ii^{\text{th}}$ element is 1, zeros elsewhere; $\boldsymbol{H}^{ij} \in \mathcal{H}_k$, $i > j$, its $ij^{\text{th}}$ and $ji^{\text{th}}$ elements are $1/\sqrt{2}$, zeros elsewhere.*
- $\widetilde{\boldsymbol{e}}_{\boldsymbol{\Sigma}}^{ij} = \frac{1}{\sqrt{\alpha}}\boldsymbol{\Sigma}^{1/2}\widetilde{\boldsymbol{H}}^{ij}\boldsymbol{\Sigma}^{1/2}$*: $\widetilde{\boldsymbol{H}}^{ij} \in \mathcal{H}_k$, its $ij^{\text{th}}$ and $ji^{\text{th}}$ elements are $\mathrm{i}/\sqrt{2}$ and $-\mathrm{i}/\sqrt{2}$, zeros elsewhere.*

*Proof:* By definition, it suffices to check that for all $1 \le p,\ell \le 2pk$, $p \ne \ell$, $\langle e_q, e_q\rangle_{\bar\theta} = 1$ and $\langle e_q, e_\ell\rangle_{\bar\theta} = 0$, which is achieved by basic calculations. ■

Recall from section II that the $q\ell^{\text{th}}$ element of the Fisher information matrix $\boldsymbol{F}_{\bar\theta}$ on $T_{\bar\theta}\overline{\mathcal{M}}_{p,k}$ is defined as $(\boldsymbol{F}_{\bar\theta})_{q\ell} = g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(e_q,e_\ell)$, where $\{e_q\}$ is given in proposition 7 and $g^{\overline{\mathcal{M}}_{p,k}}$ is defined in corollary 1. Notice that due to the invariance with respect to the action of unitary matrices in $\mathcal{U}_k$ described in section III-B, $\boldsymbol{F}_{\bar\theta}$, whose size is $2pk \times 2pk$, has rank $2pk - k^2$. This Fisher information matrix admits a particular structure, which is presented in proposition 8.

**Proposition 8.** *The Fisher information matrix $\boldsymbol{F}_{\bar\theta}$ on $\overline{\mathcal{M}}_{p,k}$ of the pdf (9) admits the structure*

$$\boldsymbol{F}_{\bar\theta} = \begin{pmatrix} \boldsymbol{F}_{\boldsymbol{U}_\perp} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F}_{\boldsymbol{U}} & \boldsymbol{F}_{\boldsymbol{U},\boldsymbol{\Sigma}} \\ \boldsymbol{0} & \boldsymbol{F}_{\boldsymbol{\Sigma},\boldsymbol{U}} & \boldsymbol{F}_{\boldsymbol{\Sigma}} \end{pmatrix},$$

*where $\boldsymbol{F}_{\boldsymbol{U}_\perp} \in \mathbb{R}^{2(p-k)k\times 2(p-k)k}$ is the block obtained from the elements $\{(\boldsymbol{e}_{\boldsymbol{U}_\perp}^{ij},\boldsymbol{0}),(\widetilde{\boldsymbol{e}}_{\boldsymbol{U}_\perp}^{ij},\boldsymbol{0})\}$ of the orthonormal basis of $T_{\bar\theta}\overline{\mathcal{M}}_{p,k}$ given in proposition 7; and $\boldsymbol{F}_{\boldsymbol{U}}$, $\boldsymbol{F}_{\boldsymbol{U},\boldsymbol{\Sigma}}$, $\boldsymbol{F}_{\boldsymbol{\Sigma},\boldsymbol{U}}$, $\boldsymbol{F}_{\boldsymbol{\Sigma}} \in \mathbb{R}^{k^2\times k^2}$ are the blocks obtained from the remaining elements of the basis. Further notice that $\boldsymbol{F}_{\boldsymbol{U}_\perp} \in \mathbb{R}^{2(p-k)k\times 2(p-k)k}$, $\boldsymbol{F}_{\boldsymbol{U}} \in \mathbb{R}^{k^2\times k^2}$ and $\boldsymbol{F}_{\boldsymbol{\Sigma}} \in \mathbb{R}^{k^2\times k^2}$ are of full rank, and*

$$\begin{pmatrix} \boldsymbol{F}_{\boldsymbol{U}} & \boldsymbol{F}_{\boldsymbol{U},\boldsymbol{\Sigma}} \\ \boldsymbol{F}_{\boldsymbol{\Sigma},\boldsymbol{U}} & \boldsymbol{F}_{\boldsymbol{\Sigma}} \end{pmatrix} \in \mathbb{R}^{2k^2\times 2k^2}$$

*has rank $k^2$.*

*Proof:* Every tangent vector $\boldsymbol{\xi}_{\boldsymbol{U}} \in T_{\boldsymbol{U}}\mathrm{St}_{p,k}$ can be decomposed as $\boldsymbol{\xi}_{\boldsymbol{U}} = \boldsymbol{U}\boldsymbol{\Omega}_\xi + \boldsymbol{U}_\perp \boldsymbol{K}_\xi$, where $\boldsymbol{U}_\perp \in \mathrm{St}_{p,p-k}$ such that $\boldsymbol{U}^H\boldsymbol{U}_\perp = \boldsymbol{0}$, $\boldsymbol{\Omega}_\xi \in \mathcal{H}_k^\perp$ and $\boldsymbol{K}_\xi \in \mathbb{C}^{(p-k)\times k}$. Thus, $\bar{\xi} \in T_{\bar\theta}\overline{\mathcal{M}}_{p,k}$ can be decomposed as

$$\bar\xi = \bar\xi^{\boldsymbol{U}} + \bar\xi^{\boldsymbol{U}_\perp} + \bar\xi^{\boldsymbol{\Sigma}} = (\boldsymbol{U}\boldsymbol{\Omega}_\xi, \boldsymbol{0}) + (\boldsymbol{U}_\perp \boldsymbol{K}_\xi, \boldsymbol{0}) + (\boldsymbol{0}, \boldsymbol{\xi}_{\boldsymbol{\Sigma}}).$$

By linearity of $g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}$ defined in corollary 1, we have

$$\begin{aligned} g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi,\bar\xi) \;=\;& g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{U}_\perp},\bar\xi^{\boldsymbol{U}_\perp}) + g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{U}},\bar\xi^{\boldsymbol{U}}) \\ &+ g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{\Sigma}},\bar\xi^{\boldsymbol{\Sigma}}) + 2g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{U}},\bar\xi^{\boldsymbol{\Sigma}}) \\ &+ 2g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{U}_\perp},\bar\xi^{\boldsymbol{U}}) + 2g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{U}_\perp},\bar\xi^{\boldsymbol{\Sigma}}). \end{aligned}$$

To show that $\boldsymbol{F}_{\bar\theta}$ has the proposed form, it suffices to prove that $g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{U}_\perp},\bar\xi^{\boldsymbol{U}}) = g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{U}_\perp},\bar\xi^{\boldsymbol{\Sigma}}) = 0$. From (17), we obtain

$$\begin{aligned} \mathrm{D}\,\bar\varphi(\bar\theta)[\bar\xi^{\boldsymbol{U}_\perp}] &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{K}_\xi^H\boldsymbol{U}_\perp^H + \boldsymbol{U}_\perp\boldsymbol{K}_\xi\boldsymbol{\Sigma}\boldsymbol{U}^H, \\ \mathrm{D}\,\bar\varphi(\bar\theta)[\bar\xi^{\boldsymbol{U}}] &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Omega}_\xi^H\boldsymbol{U}^H + \boldsymbol{U}\boldsymbol{\Omega}_\xi\boldsymbol{\Sigma}\boldsymbol{U}^H, \\ \mathrm{D}\,\bar\varphi(\bar\theta)[\bar\xi^{\boldsymbol{\Sigma}}] &= \boldsymbol{U}\boldsymbol{\xi}_{\boldsymbol{\Sigma}}\boldsymbol{U}^H. \end{aligned}$$

The Woodbury identity $\bar\varphi(\bar\theta)^{-1} = \boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{\Xi}\boldsymbol{U}^H$, where $\boldsymbol{\Xi} = (\boldsymbol{I}_k + \boldsymbol{\Sigma}^{-1})^{-1}$, and $\boldsymbol{U}^H\boldsymbol{U}_\perp = \boldsymbol{0}$ leads to

$$\begin{aligned} \bar\varphi(\bar\theta)^{-1}\,\mathrm{D}\,\bar\varphi(\bar\theta)[\bar\xi^{\boldsymbol{U}_\perp}] = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{K}_\xi^H\boldsymbol{U}_\perp^H + \boldsymbol{U}_\perp\boldsymbol{K}_\xi\boldsymbol{\Sigma}\boldsymbol{U}^H \\ - \boldsymbol{U}\boldsymbol{\Xi}\boldsymbol{\Sigma}\boldsymbol{K}_\xi^H\boldsymbol{U}_\perp^H, \end{aligned}$$

from which one can check that $\mathrm{tr}(\bar\varphi(\bar\theta)^{-1}\,\mathrm{D}\,\bar\varphi(\bar\theta)[\bar\xi^{\boldsymbol{U}_\perp}]) = 0$. Furthermore, the previous expression yields

$$\begin{aligned} \bar\varphi(\bar\theta)^{-1}\,\mathrm{D}\,\bar\varphi(\bar\theta)[\bar\xi^{\boldsymbol{U}_\perp}]\bar\varphi(\bar\theta)^{-1} = \boldsymbol{U}_\perp\boldsymbol{K}_\xi\boldsymbol{\Sigma}(\boldsymbol{I}_k - \boldsymbol{\Xi})\boldsymbol{U}^H \\ + \boldsymbol{U}(\boldsymbol{I}_k - \boldsymbol{\Xi})\boldsymbol{\Sigma}\boldsymbol{K}_\xi^H\boldsymbol{U}_\perp^H. \end{aligned}$$

From this, it is readily checked that $g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{U}_\perp},\bar\xi^{\boldsymbol{U}}) = g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{U}_\perp},\bar\xi^{\boldsymbol{\Sigma}}) = 0$. Finally, to show that $\boldsymbol{F}_{\boldsymbol{U}_\perp} \in \mathbb{R}^{2(p-k)k\times 2(p-k)k}$, $\boldsymbol{F}_{\boldsymbol{U}} \in \mathbb{R}^{k^2\times k^2}$ and $\boldsymbol{F}_{\boldsymbol{\Sigma}} \in \mathbb{R}^{k^2\times k^2}$ are of full rank, it is enough to verify that $\bar\xi^{\boldsymbol{U}_\perp} \mapsto g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{U}_\perp},\bar\xi^{\boldsymbol{U}_\perp})$, $\bar\xi^{\boldsymbol{U}} \mapsto g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{U}},\bar\xi^{\boldsymbol{U}})$ and $\bar\xi^{\boldsymbol{\Sigma}} \mapsto g_{\bar\theta}^{\overline{\mathcal{M}}_{p,k}}(\bar\xi^{\boldsymbol{\Sigma}},\bar\xi^{\boldsymbol{\Sigma}})$ are positive definite. The rank of

$$\begin{pmatrix} \boldsymbol{F}_{\boldsymbol{U}} & \boldsymbol{F}_{\boldsymbol{U},\boldsymbol{\Sigma}} \\ \boldsymbol{F}_{\boldsymbol{\Sigma},\boldsymbol{U}} & \boldsymbol{F}_{\boldsymbol{\Sigma}} \end{pmatrix} \in \mathbb{R}^{2k^2\times 2k^2}$$

is given by subtracting the rank of $\boldsymbol{F}_{\boldsymbol{U}_\perp}$ to the one of $\boldsymbol{F}_{\bar\theta}$. ■

Also recall from section II that, given an unbiased estimator $\widehat\theta$ of $\theta$ in $\mathcal{M}_{p,k}$, the error matrix is $\boldsymbol{C}_{\bar\theta} = \boldsymbol{x}_{\bar\theta}\boldsymbol{x}_{\bar\theta}^T$, where the $q^{\text{th}}$ element of $\boldsymbol{x}_{\bar\theta}$ is $(\boldsymbol{x}_{\bar\theta})_q = \langle\overline{\log_\theta(\widehat\theta)}, e_q\rangle_{\bar\theta}$. However, as discussed in section IV, the Riemannian logarithm map (as well as the distance) on $\mathcal{M}_{p,k}$ is not known in closed form. Hence, the corresponding intrinsic Cramér-Rao bound inequality only offers a theoretical ideal that cannot easily be measured in practice. This issue motivates the derivation of a new estimation error measure on $\mathcal{M}_{p,k}$, as well as the associated performance bound.

### *C. Alternative intrinsic Cramér-Rao bound associated to a non-Riemannian geometry on $\mathcal{M}_{p,k}$*

As the Riemannian logarithm map and distance are unknown on $\mathcal{M}_{p,k}$, we need to define new alternative geometrical objects on $\mathcal{M}_{p,k}$ in order to be able to measure the estimation error and accurately bound it. To do so, inspired by [23], we consider the alternative horizontal space at $\bar\theta \in \overline{\mathcal{M}}_{p,k}$

$$\widetilde{\mathcal{H}}_{\bar\theta} = \{\bar\xi \in T_{\bar\theta}\overline{\mathcal{M}}_{p,k} : \boldsymbol{U}^H\boldsymbol{\xi}_{\boldsymbol{U}} = \boldsymbol{0}\}. \tag{20}$$

$\widetilde{\mathcal{H}}_{\bar\theta}$ still provides proper representatives of the elements in $T_\theta\mathcal{M}_{p,k}$, *i.e.*, there is a one to one correspondance between elements in $\widetilde{\mathcal{H}}_{\bar\theta}$ and vectors in $T_\theta\mathcal{M}_{p,k}$: given $\xi \in T_\theta\mathcal{M}_{p,k}$, there is a unique $\bar\xi \in \widetilde{\mathcal{H}}_{\bar\theta}$ such that $\xi = \mathrm{D}\,\pi(\bar\theta)[\bar\xi]$. This horizontal space is advantageous because the geodesics in $\overline{\mathcal{M}}_{p,k}$ emanating from it are well characterized: the part of the geodesics that concerns $\boldsymbol{U}$ coincides with the geodesics of the Grassmann manifold $\mathcal{G}_{p,k}$ while the part that concerns $\boldsymbol{\Sigma}$ does not change. These yield proper curves $\widetilde\gamma$ in $\mathcal{M}_{p,k}$, which allow to join any two points $\theta = \pi(\boldsymbol{U},\boldsymbol{\Sigma})$ and $\vartheta = \pi(\boldsymbol{V},\boldsymbol{\Gamma})$. The curve $\widetilde\gamma : [0,1] \to \mathcal{M}_{p,k}$, with $\widetilde\gamma(0) = \theta$ and $\widetilde\gamma(1) = \vartheta$, is $\widetilde\gamma(t) = \pi(\widetilde{\boldsymbol{U}}(t), \widetilde{\boldsymbol{\Sigma}}(t))$ such that $(\widetilde{\boldsymbol{U}}(t), \widetilde{\boldsymbol{\Sigma}}(t))$ is the geodesic on $\overline{\mathcal{M}}_{p,k}$ defined as

$$\begin{aligned} \widetilde{\boldsymbol{U}}(t) &= \boldsymbol{U}\boldsymbol{O}\cos(t\boldsymbol{\Theta})\boldsymbol{O}^H \\ &\quad + (\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}^H)\boldsymbol{V}\boldsymbol{P}(\sin(\boldsymbol{\Theta}))^\dagger\sin(t\boldsymbol{\Theta})\boldsymbol{O}^H, \\ \widetilde{\boldsymbol{\Sigma}}(t) &= \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{O}\boldsymbol{P}^H\boldsymbol{\Gamma}\boldsymbol{P}\boldsymbol{O}^H\boldsymbol{\Sigma}^{-1/2})^t\boldsymbol{\Sigma}^{1/2}, \end{aligned} \tag{21}$$

where $\boldsymbol{O}$, $\boldsymbol{P}$ and $\boldsymbol{\Theta}$ correspond to the singular value decomposition $\boldsymbol{U}^H\boldsymbol{V} = \boldsymbol{O}\cos(\boldsymbol{\Theta})\boldsymbol{P}^H$, $\cdot^\dagger$ and $\cdot^t = \exp(t\log(\cdot))$ are Moore-Penrose pseudo-inverse and matrix power functions, respectively.

However, since $\widetilde{\mathcal{H}}_{\bar\theta}$ is not the orthogonal complement to the vertical space $\mathcal{V}_{\bar\theta}$ according to the chosen metric $\langle\cdot,\cdot\rangle_{\bar\theta}$, these curves are not geodesics in $\mathcal{M}_{p,k}$. Yet, it remains possible to define the associated "logarithm" map and divergence on $\mathcal{M}_{p,k}$, which are given in proposition 9. Given $\theta$ and $\vartheta$, the "logarithm" of $\vartheta$ at $\theta$ is the tangent vector in $T_\theta\mathcal{M}_{p,k}$ represented by the element in $\widetilde{\mathcal{H}}_{\bar\theta}$ corresponding to the curve (21).

**Proposition 9.** *The "logarithm" $\widetilde{\log}_\theta(\vartheta)$ of $\vartheta = \pi(\boldsymbol{V},\boldsymbol{\Gamma})$ at $\theta = \pi(\boldsymbol{U},\boldsymbol{\Sigma})$ associated with curve* (21) *is the tangent vector in $T_\theta\mathcal{M}_{p,k}$ represented in $\widetilde{\mathcal{H}}_{\bar\theta}$ by*

$$\begin{aligned} \overline{\widetilde{\log}_\theta(\vartheta)} = (&\boldsymbol{X}\boldsymbol{\Theta}\boldsymbol{Y}^H, \\ &\boldsymbol{\Sigma}^{1/2}\log(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{O}\boldsymbol{P}^H\boldsymbol{\Gamma}\boldsymbol{P}\boldsymbol{O}^H\boldsymbol{\Sigma}^{-1/2})\boldsymbol{\Sigma}^{1/2}), \end{aligned}$$

*where one has the two singular value decompositions $\boldsymbol{U}^H\boldsymbol{V} = \boldsymbol{O}\cos(\boldsymbol{\Theta})\boldsymbol{P}^H$ and $(\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}^H)\boldsymbol{V}(\boldsymbol{U}^H\boldsymbol{V})^{-1} = \boldsymbol{X}\tan(\boldsymbol{\Theta})\boldsymbol{Y}^H$. Furthermore, measuring the squared length of the curve* (21) *yield the divergence function on $\mathcal{M}_{p,k}$*

$$\begin{aligned} d_{\mathcal{M}_{p,k}}(\theta,\vartheta) = \alpha\left\|\log(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{O}\boldsymbol{P}^H\boldsymbol{\Gamma}\boldsymbol{P}\boldsymbol{O}^H\boldsymbol{\Sigma}^{-1/2})\right\|_F^2 \\ + \beta(\log\det(\boldsymbol{\Sigma}^{-1}\boldsymbol{O}\boldsymbol{P}^H\boldsymbol{\Gamma}\boldsymbol{P}\boldsymbol{O}^H))^2 + \|\boldsymbol{\Theta}\|_F^2. \end{aligned}$$

*Proof.* Concerning the "logarithm" map, curve (21) is composed of a geodesic on $\mathcal{G}_{p,k}$ and one on $\mathcal{H}_k^{++}$. To conclude, it is thus enough to notice that the proposed "logarithm" is defined through the logarithm maps of $\mathcal{G}_{p,k}$ (see *e.g.* [18], [48]) and of $\mathcal{H}_k^{++}$ (see *e.g.* [47]).

The divergence is obtained by measuring the squared length of curve (21), which is the sum of the squared distances on $\mathcal{G}_{p,k}$ and $\mathcal{H}_k^{++}$. Notice that for $\alpha = 1$ and $\beta = 0$, we obtain the divergence in [23]. ∎

Similarly to the usual intrinsic Cramér-Rao bound, an orthonormal basis of $\widetilde{\mathcal{H}}_{\bar{\theta}}$ is required in order to define the Fisher information matrix and the estimation error matrix in our case. From the definition (20) of $\widetilde{\mathcal{H}}_{\bar{\theta}}$ and of the chosen metric (12), one can check that such basis $\{\widetilde{e}_q\}_{1\leq q\leq 2pk-k^2}$ is given by

$$\left\{ \{(\boldsymbol{e}_{\boldsymbol{U}_\perp}^{ij}, \boldsymbol{0}), (\widetilde{\boldsymbol{e}}_{\boldsymbol{U}_\perp}^{ij}, \boldsymbol{0})\}_{\substack{1\leq i\leq p-k,\\ 1\leq j\leq k}}, \right. \\ \left. \{(\boldsymbol{0}, \boldsymbol{e}_{\boldsymbol{\Sigma}}^{ij})\}_{1\leq j\leq i\leq k}, \{(\boldsymbol{0}, \widetilde{\boldsymbol{e}}_{\boldsymbol{\Sigma}}^{ij})\}_{1\leq j< i\leq k} \right\},$$

where all these elements are defined in proposition 8. Concerning the Fisher information matrix $\widetilde{\boldsymbol{F}}_{\bar{\theta}}$ on $\widetilde{\mathcal{H}}_{\bar{\theta}}$, its $q\ell^{\text{th}}$ element is defined as $(\widetilde{\boldsymbol{F}}_{\bar{\theta}})_{q\ell} = g_{\bar{\theta}}^{\overline{\mathcal{M}}_{p,k}}(\widetilde{e}_q, \widetilde{e}_\ell)$, where $g^{\overline{\mathcal{M}}_{p,k}}$ is defined in corollary 1. It follows that $\widetilde{\boldsymbol{F}}_{\bar{\theta}}$ admits the structure

$$\widetilde{\boldsymbol{F}}_{\bar{\theta}} = \begin{pmatrix} \boldsymbol{F}_{\boldsymbol{U}_\perp} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F}_{\boldsymbol{\Sigma}} \end{pmatrix}, \tag{22}$$

where $\boldsymbol{F}_{\boldsymbol{U}_\perp}$ and $\boldsymbol{F}_{\boldsymbol{\Sigma}}$ are defined in proposition 8. Furthermore, its size is $(2pk - k^2) \times (2pk - k^2)$ and it has full rank. Interestingly, $\widetilde{\boldsymbol{F}}_{\bar{\theta}}$ corresponds to the Fisher information matrix obtained in [18] from a different reasonning for the Gaussian case ($\alpha^{++} = 1$ in corollary 1) and for $\alpha = 1$ and $\beta = 0$ in metric (12).

Given an estimator $\widehat{\theta}$ of $\theta \in \mathcal{M}_{p,k}$, the estimation error matrix is $\widetilde{\boldsymbol{C}}_{\bar{\theta}} = \widetilde{\boldsymbol{x}}_{\bar{\theta}}\widetilde{\boldsymbol{x}}_{\bar{\theta}}^T$, such that the $q^{\text{th}}$ element of $\widetilde{\boldsymbol{x}}_{\bar{\theta}}$ is $(\widetilde{\boldsymbol{x}}_{\bar{\theta}})_q = \langle \widetilde{\overline{\log}_\theta(\widehat{\theta})}, \widetilde{e}_q\rangle_{\bar{\theta}}$, where $\widetilde{\overline{\log}_\theta(\widehat{\theta})} \in \widetilde{\mathcal{H}}_{\bar{\theta}}$ is defined in proposition 9. From these alternative Fisher information and estimation error matrices, it is possible to define an alternative intrinsic Cramér-Rao bound on $\mathcal{M}_{p,k}$ for complex elliptically symmetric distributions with covariance matrix of the form (4). It is achieved in proposition 10.

**Proposition 10.** *The error estimation matrix $\widetilde{\boldsymbol{C}}_{\bar{\theta}}$ on $\widetilde{\mathcal{H}}_{\bar{\theta}}$ of an unbiased estimator $\widehat{\theta}$ of $\theta$ in $\mathcal{M}_{p,k}$ admits the lower bound*

$$\mathbb{E}[\widetilde{\boldsymbol{C}}_{\bar{\theta}}] \succeq \widetilde{\boldsymbol{F}}_{\bar{\theta}}^{-1} + \textit{curvature terms},$$

*where, as in [18], curvature terms*[7] *can be neglected at small errors. Moreover, taking the trace of this inequality yields*

$$\mathbb{E}[\mathrm{err}_\theta^{\mathcal{M}_{p,k}}(\widehat{\theta})] \geq \mathrm{tr}(\widetilde{\boldsymbol{F}}_{\bar{\theta}}^{-1}) = \mathrm{tr}(\boldsymbol{F}_{\boldsymbol{U}_\perp}^{-1}) + \mathrm{tr}(\boldsymbol{F}_{\boldsymbol{\Sigma}}^{-1}),$$

*where* $\mathrm{err}_\theta^{\mathcal{M}_{p,k}}(\widehat{\theta}) = d_{\mathcal{M}_{p,k}}(\theta, \widehat{\theta})$.

*Proof.* Let $\widetilde{\boldsymbol{x}}_{\bar{\theta}}$ and $\widetilde{\boldsymbol{s}}_{\bar{\theta}}$ such that their $q^{\text{th}}$ elements are $(\widetilde{\boldsymbol{x}}_{\bar{\theta}})_q = \langle \widetilde{\overline{\log}_\theta(\widehat{\theta})}, \widetilde{e}_q\rangle_{\bar{\theta}}$ and $(\widetilde{\boldsymbol{s}}_{\bar{\theta}})_q = \mathrm{D}\,\overline{L}_g(\bar{\theta})[\widetilde{e}_q]$. As $\widetilde{\overline{\log}_\theta(\widehat{\theta})}$ is composed of the Riemannian logarithm maps of $\mathcal{G}_{p,k}$ and $\mathcal{H}_k^{++}$, [18, Lemma 1] remains valid in this case and one has

$$\mathbb{E}[\widetilde{\boldsymbol{x}}_{\bar{\theta}}\widetilde{\boldsymbol{s}}_{\bar{\theta}}^T] = \boldsymbol{I}_{2pk-k^2} - \frac{1}{3}\widetilde{R}_m(\widetilde{\boldsymbol{C}}_{\bar{\theta}}) + O(\|\widetilde{\boldsymbol{x}}_{\bar{\theta}}\|^3),$$

where $\widetilde{R}_m(\widetilde{\boldsymbol{C}}_{\bar{\theta}})$ is related to the curvature of $\mathcal{M}_{p,k}$ associated with the considered geometry, *i.e.*, the one resulting from the alternative horizontal space $\widetilde{\mathcal{H}}_{\bar{\theta}}$. Curvature terms can be computed by following and adapting the main lines of [18] to this context. However, as they can be neglected at small errors, for simplicity and clarity, we do not explicit $\widetilde{R}_m$ more in the following. From there, [18, Theorem 2] can also be adapted. Let $\boldsymbol{v} = \widetilde{\boldsymbol{x}}_{\bar{\theta}} - \widetilde{\boldsymbol{F}}_{\bar{\theta}}^{-1}\widetilde{\boldsymbol{s}}_{\bar{\theta}}$. Expanding $\mathbb{E}[\boldsymbol{v}\boldsymbol{v}^T]$, noticing that it is a positive semi-definite matrix and that $\mathbb{E}[\widetilde{\boldsymbol{s}}_{\bar{\theta}}\widetilde{\boldsymbol{s}}_{\bar{\theta}}^T] = \widetilde{\boldsymbol{F}}_{\bar{\theta}}$ yields

$$\mathbb{E}[\widetilde{\boldsymbol{C}}_{\bar{\theta}}] \succeq \widetilde{\boldsymbol{F}}_{\bar{\theta}}^{-1} + \text{curvature terms},$$

which is enough to conclude. ∎

The new error measure $\mathrm{err}^{\mathcal{M}_{p,k}}$ of proposition 10 can also be bounded with the bound of the previous section. Indeed, the divergence $d_{\mathcal{M}_{p,k}}$ on $\mathcal{M}_{p,k}$ is obtained by measuring the length according to $\langle\cdot,\cdot\rangle_\cdot$ of a non-minimal curve in $\mathcal{M}_{p,k}$. Hence, we have the inequality $d_{\mathcal{M}_{p,k}}(\theta,\widehat{\theta}) \geq \delta^2_{\mathcal{M}_{p,k}}(\theta,\widehat{\theta})$, which yields the lower bound

$$\mathbb{E}[\mathrm{err}_\theta^{\mathcal{M}_{p,k}}(\widehat{\theta})] \geq \mathrm{tr}(\boldsymbol{F}_{\bar{\theta}}^{\dagger}), \tag{23}$$

However, by construction, one might expect this lower bound not to be tight.

### D. Intrinsic Cramér-Rao bound on $\mathcal{G}_{p,k}$ for subspace recovery

The principal subspace of $\varphi(\theta) \in \mathcal{H}_{p,k}^{+}$ is given by $\mathrm{span}(\boldsymbol{U})$ in the Grassmann manifold $\mathcal{G}_{p,k}$. Thanks to the structure of the alternative Cramér-Rao bound obtained in the previous section, we can provide an intrinsic Cramér-Rao bound on $\mathcal{G}_{p,k}$ for a given estimate of this subspace, denoted $\mathrm{span}(\widehat{\boldsymbol{U}})$. Indeed, both $\widetilde{\boldsymbol{C}}_{\bar{\theta}}$ and $\widetilde{\boldsymbol{F}}_{\bar{\theta}}$ are block diagonal matrices. Let $\boldsymbol{C}_{\boldsymbol{U}_\perp}$ be the block of $\widetilde{\boldsymbol{C}}_{\bar{\theta}}$ that concerns $\boldsymbol{U}$. Both $\boldsymbol{C}_{\boldsymbol{U}_\perp}$ and $\boldsymbol{F}_{\boldsymbol{U}_\perp}$ are by definition on the horizontal space of $\mathcal{G}_{p,k}$, *i.e.*, they are constructed with an orthonormal basis on the horizontal space and the logarithm map on $\mathcal{G}_{p,k}$. The corresponding intrinsic Cramér-Rao bound on $\mathcal{G}_{p,k}$ along with a closed form formula for $\mathrm{tr}(\boldsymbol{F}_{\boldsymbol{U}_\perp}^{-1})$ are provided in proposition 11.

**Proposition 11.** *The subspace estimation error $\boldsymbol{C}_{\boldsymbol{U}_\perp}$ admits the lower bound*

$$\mathbb{E}[\boldsymbol{C}_{\boldsymbol{U}_\perp}] \succeq \boldsymbol{F}_{\boldsymbol{U}_\perp}^{-1}.$$

*Taking the trace of the inequality yields*

$$\mathbb{E}[\mathrm{err}_\theta^{\mathcal{G}_{p,k}}(\widehat{\theta})] \geq \mathrm{tr}(\boldsymbol{F}_{\boldsymbol{U}_\perp}^{-1}) = \frac{(p-k)}{n\alpha^{++}}\sum_{i=1}^k \frac{1+\sigma_i}{\sigma_i^2},$$

*where $\{\sigma_i\}$ are the eigenvalues of $\boldsymbol{\Sigma}$ and $\mathrm{err}_\theta^{\mathcal{G}_{p,k}}(\widehat{\theta}) = \delta^2_{\mathcal{G}_{p,k}}(\mathrm{span}(\boldsymbol{U}), \mathrm{span}(\widehat{\boldsymbol{U}}))$. The Riemannian distance on $\mathcal{G}_{p,k}$ is $\delta^2_{\mathcal{G}_{p,k}}(\mathrm{span}(\boldsymbol{U}), \mathrm{span}(\widehat{\boldsymbol{U}})) = \|\boldsymbol{\Theta}\|_F^2$, where $\boldsymbol{\Theta}$ is defined in proposition 9.*

*Proof.* The first inequality is straightforward from proposition 10 and from the fact that $\widetilde{\boldsymbol{C}}_{\bar{\theta}}$ and $\widetilde{\boldsymbol{F}}_{\bar{\theta}}$ are block diagonal.

[7] Notice that these curvature terms are different from those in the intrinsic Cramér-Rao bound associated with the Riemannian geometry of $\mathcal{M}_{p,k}$ presented in section IV. Indeed, an alternative horizontal space yielding different curves on $\mathcal{M}_{p,k}$ is considered.

Thus, it only remains to show $\operatorname{tr}(\boldsymbol{F}_{\boldsymbol{U}_\perp}^{-1}) = \frac{(p-k)}{n\alpha^{++}} \sum_{i=1}^{k} \frac{1+\sigma_i}{\sigma_i^2}$. Given $\bar{\xi}^{\boldsymbol{U}_\perp}$ and $\bar{\eta}^{\boldsymbol{U}_\perp}$ defined as in the proof of proposition 8, one can check that

$$g_{\bar{\theta}}^{\overline{\mathcal{M}}_{p,k}}(\bar{\xi}^{\boldsymbol{U}_\perp}, \bar{\eta}^{\boldsymbol{U}_\perp}) \\ = 2n\alpha^{++} \operatorname{tr}(\boldsymbol{\Sigma}(\boldsymbol{I} - \boldsymbol{\Xi})\boldsymbol{\Sigma} \operatorname{herm}(\boldsymbol{K}_\xi \boldsymbol{K}_\eta)),$$

where $\boldsymbol{\Xi} = (\boldsymbol{I}_k + \boldsymbol{\Sigma}^{-1})^{-1}$ as before. Let us choose the representative in equivalence class (8) such that $\boldsymbol{\Sigma}$ is diagonal (with elements $\{\sigma_i\}$). It follows that $\boldsymbol{\Sigma}(\boldsymbol{I} - \boldsymbol{\Xi})\boldsymbol{\Sigma}$ is the diagonal matrix whose elements are $\{\frac{\sigma_i^2}{1+\sigma_i}\}$. To compute $\boldsymbol{F}_{\boldsymbol{U}_\perp}$, elements $\boldsymbol{K}_\xi$ and $\boldsymbol{K}_\eta$ are taken from matrices $\boldsymbol{K}^{ij}$ defined in proposition 7. With proper ordering of the basis elements, one obtains

$$\boldsymbol{F}_{\boldsymbol{U}_\perp} = 2n\alpha^{++} \boldsymbol{I}_{2(p-k)} \otimes (\boldsymbol{\Sigma}(\boldsymbol{I} - \boldsymbol{\Xi})\boldsymbol{\Sigma}).$$

Taking the trace of $\boldsymbol{F}_{\boldsymbol{U}_\perp}^{-1}$ yields the proposed result. To verify that this remains valid when $\boldsymbol{\Sigma}$ is not diagonal, it is enough to notice that both sides of the equality are invariant with respect to the action of $\mathcal{U}_k$ in the equivalence class. ■

## VII. Numerical experiments

### A. Validation simulations

This section illustrates our Riemannian optimization framework and performance analysis for robust covariance estimation. In order to do so, we perform covariance estimation of simulated data drawn from the multivariate Student $t$-distribution with $d = 3$ (highly non-Gaussian) and $d = 100$ (almost Gaussian) degrees of freedom; see [1] for details.

To generate a covariance matrix admitting the structure (4), we compute

$$\boldsymbol{R} = \boldsymbol{I}_p + \sigma \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H,$$

where

- $\boldsymbol{U}$ is a random matrix in $\mathrm{St}_{p,k}$,
- $\boldsymbol{\Sigma}$ is a diagonal matrix whose minimal and maximal elements are $1/\sqrt{c}$ and $\sqrt{c}$ ($c = 20$ is the condition number with respect to inversion of $\boldsymbol{\Sigma}$); its other elements are randomly drawn from the uniform distribution between $1/\sqrt{c}$ and $\sqrt{c}$; its trace is then normalized as $\operatorname{tr}(\boldsymbol{\Sigma}) = \operatorname{tr}(\boldsymbol{I}_k) = k$,
- $\sigma = 50$ is a free parameter corresponding to the spike to noise ratio.

In our experiment, we choose $p = 16$ and $k \in \{4, 8\}$. Sets $\{\boldsymbol{x}_i\}_{i=1}^n$ are drawn from the multivariate Student $t$-distribution with covariance $\boldsymbol{R}$ and $d \in \{3, 100\}$, where $n \in \{12, 14, 15, 17, 20, 40, 70, 100, 200, 300\}$. For each value of $n$, 500 sets $\{\boldsymbol{x}_i\}_{i=1}^n$ are simulated and the aim is to estimate the structured covariance matrix $\boldsymbol{R}$ in each case.

The considered estimators in this experiment are:

(a) Projected sample covariance matrix $\boldsymbol{I}_p + \varphi(\widehat{\theta}_{\text{pSCM}})$ obtained by projecting $n^{-1}\sum_i \boldsymbol{x}_i\boldsymbol{x}_i^H$ on $\boldsymbol{I}_p + \mathcal{H}_{p,k}^+$ with [32, equation (53)].
(b) Structured Tyler's $M$-estimator $\boldsymbol{I}_p + \varphi(\widehat{\theta}_{\text{T-MM}})$ solved with [32, algorithm 5].
(c) Structured Tyler's $M$-estimator $\boldsymbol{I}_p + \varphi(\widehat{\theta}_{\text{T-RGD}})$ solved with a Riemannian gradient descent algorithm on $\mathcal{M}_{p,k}$; see [33, chapter 4].
(d) Structured Tyler's $M$-estimator $\boldsymbol{I}_p + \varphi(\widehat{\theta}_{\text{T-RTR}})$ solved with a Riemannian trust region algorithm (second order optimization method) on $\mathcal{M}_{p,k}$; see [33, chapter 7].

The three iterative methods are initialized with the principal subspace of the projected sample covariance matrix estimator, *i.e.*, $(\widehat{\boldsymbol{U}}_{\text{pSCM}}, \boldsymbol{I}_k)$. Riemannian optimization on $\mathcal{M}_{p,k}$ is performed with manopt toolbox [49] and we choose $\alpha = \frac{p+d}{p+d+1}$ and $\beta = \alpha - 1$ in the Riemannian metric (12).

In figures 2 and 3, we observe that, in all considered cases, *i.e.* $d \in \{3, 100\}$ and $k \in \{4, 8\}$, the lower bound (23) is not reached by any of the methods for error measure of proposition 10. This is expected as this bound is suited to the Riemannian distance on $\mathcal{M}_{p,k}$ and not to the divergence of proposition 9. However, the bound of proposition 10, which arises from the Fisher information matrix well suited to our divergence, is reached by several methods as the number of samples $n$ grows. Concerning the subspace error, the lower bound in proposition 11 is reached in all considered cases by several methods as $n$ grows. Further notice that, for $k = 4$, a smaller amount of samples $n$ is needed for the bounds of propositions 10 and 11 to be attained than for $k = 8$.

Unlike the other considered estimators, the performance of pSCM depends on the degree of freedom $d$ of the Student $t$-distribution. As expected, when data are close to Gaussianity ($d = 100$), pSCM provides good results and attains both bounds of propositions 10 and 11. However, when they are far from being Gaussian ($d = 3$), pSCM fails to give optimal results. We also observe that T-MM and T-RTR have very similar performance. They both fail when $n$ is small, especially when it gets close to $p$ (or smaller). However, they perform well when $n$ is sufficient and reach both bounds of propositions 10 and 11. Concerning T-RGD, we notice that it yields good results as compared to other estimators when $n$ is small. As $n$ grows, even though T-RGD still provide satisfying subspaces (it reaches the bound of proposition 11), its performance with respect to error measure of proposition 10 deteriorates as compared to other estimators. In conclusion, our optimization framework on $\mathcal{M}_{p,k}$ provides satisfying results on these simulated data for all considered cases. Depending on the number of samples at hand, different optimization algorithms are preferable: the first order method (T-RGD) is more advantageous when a small amount of samples is available whereas the second order method (T-RTR) performs better as the number of samples grows.

### B. Application to adaptive filtering

The optimal filter in terms of signal to noise ratio is built as $\boldsymbol{w} = \boldsymbol{R}^{-1}\boldsymbol{p}$, where $\boldsymbol{R}$ is the interference covariance matrix, which is of the form (4), and $\boldsymbol{p}$ is the steering vector. In practice the interference covariance matrix is unknown and $n$ signal-free samples $\{\boldsymbol{x}_i\}_{i=1}^n$ are used in order to build a so-called adaptive filter as $\widehat{\boldsymbol{w}} = \widehat{\boldsymbol{R}}^{-1}\boldsymbol{p}$, where $\widehat{\boldsymbol{R}}$ is an estimate of $\boldsymbol{R}$. In this case, the performance of the adaptive filter highly depends on the accuracy of the estimation step. In a classical

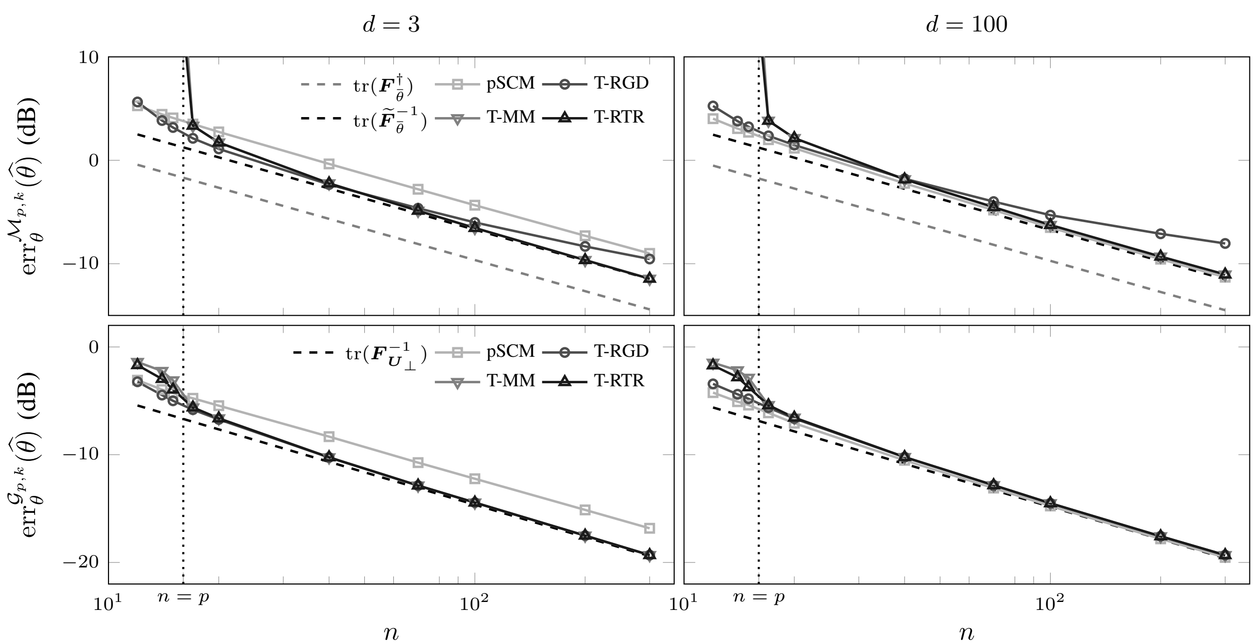


Fig. 2: Mean of error measures in propositions 10 (top) and 11 (bottom) of methods pSCM, T-MM, T-RGD and T-RTR along with corresponding intrinsic Cramér-Rao bounds in (23) and propositions 10 and 11 as functions of the number of samples $n$. Means are computed over 500 simulated sets $\{\boldsymbol{x}_i\}_{i=1}^n$ with $d = 3$ (left) and 100 (right), $p = 16$ and $k = 4$.

Gaussian setting, a $-3$dB of SNRLoss (expected loss in signal to noise ratio compared to optimal filter), defined as

$$\mathrm{SNRLoss}(\widehat{\boldsymbol{R}}) = \mathbb{E}\left[\frac{(\widehat{\boldsymbol{w}}^H\boldsymbol{p})^2}{(\widehat{\boldsymbol{w}}^H\boldsymbol{R}\widehat{\boldsymbol{w}})(\boldsymbol{p}^H\boldsymbol{R}^{-1}\boldsymbol{p})}\right] \tag{24}$$

is achieved with the sample covariance matrix for $n = 2p$.

This performance is generally degraded when the interference is not Gaussian and/or can also be improved by taking prior information on the covariance matrix structure in the estimation step. This is illustrated in Figure 4 where the SNR-Loss of the adaptive filters built from the previous estimators is evaluated through Monte-Carlo simulations under the following setting: $p = 16$, $\boldsymbol{H}$ is the rank $k = 8$ truncation of a Toeplitz matrix $(\boldsymbol{\Sigma}_{\mathrm{T}})_{i,j} = (0.9(i+1)/\sqrt{2})^{|i-j|}$, the spike to noise ratio is 20dB, samples following a Student $t$-distribution with $d \in \{3, 100\}$, Again, we observe the interest of the proposed estimation methods, notably at low sample support.

## VIII. Conclusions and perspectives

This article proposes an original Riemannian geometry to study low-rank structured elliptical models. The tools developed within this framework (representations of tangent spaces, geodesics, Riemannian gradient and Hessian, retraction, divergence function) allow to derive both estimation algorithms and intrinsic Cramér-Rao lower bounds adapted to these models with a unified view. Some potential extensions of this work include: generalization to $M$-estimators and estimation of the parameters of the Fisher information metric, integration of curvature terms and intrinsic bias in the intrinsic Cramér-Rao lower bounds, generalizations of expectation-maximization type algorithms [50] to handle multimodal mixture models [51].

## References


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

[2] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image processing*, 12(11):1338–1351, 2003.

[3] F. Shi and I. W. Selesnick. An elliptically contoured exponential mixture model for wavelet based image denoising. *Applied and Computational Harmonic Analysis*, 23(1):131–151, 2007.

[4] S. Zozor and C. Vignat. Some results on the denoising problem in the elliptically distributed context. *IEEE Transactions on Signal Processing*, 58(1):134–150, 2010.

[5] M. Greco, F. Gini, and M. Rangaswamy. Statistical analysis of measured polarimetric clutter data at different range resolutions. *IEE Proceedings-Radar, Sonar and Navigation*, 153(6):473–481, 2006.

[6] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor. Compound-Gaussian clutter modeling with an inverse Gaussian texture distribution. *IEEE Signal Processing Letters*, 19(12):876–879, 2012.

[7] R. A. Maronna and V. J. Yohai. Robust estimation of multivariate location and scatter. *Wiley StatsRef: Statistics Reference Online*, 1976.

[8] D. E. Tyler. A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, pages 234–251, 1987.

[9] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295–327, 2001.

[10] M.E. Tipping and C.M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

[11] G. Drašković, A. Breloy, and F. Pascal. On the asymptotics of maronna's robust PCA. *IEEE Transactions on Signal Processing*, 67(19):4964–4975, 2019.

[12] G. Lerman and T. Maunu. An overview of robust subspace recovery. *Proceedings of the IEEE*, 106(8):1380–1410, 2018.

[13] M. Haardt, M. Pesavento, F. Roemer, and M. N. El Korso. Chapter 15 - subspace methods and exploitation of special array structures. In A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, editors, *Academic Press Library in Signal Processing: Volume 3*, volume 3 of *Academic Press Library in Signal Processing*, pages 651–717. Elsevier, 2014.

[14] A. Haimovich. The eigencanceler: Adaptive radar by eigenanalysis methods. *IEEE Transactions on Aerospace and Electronic Systems*, 32(2):532–542, 1996.

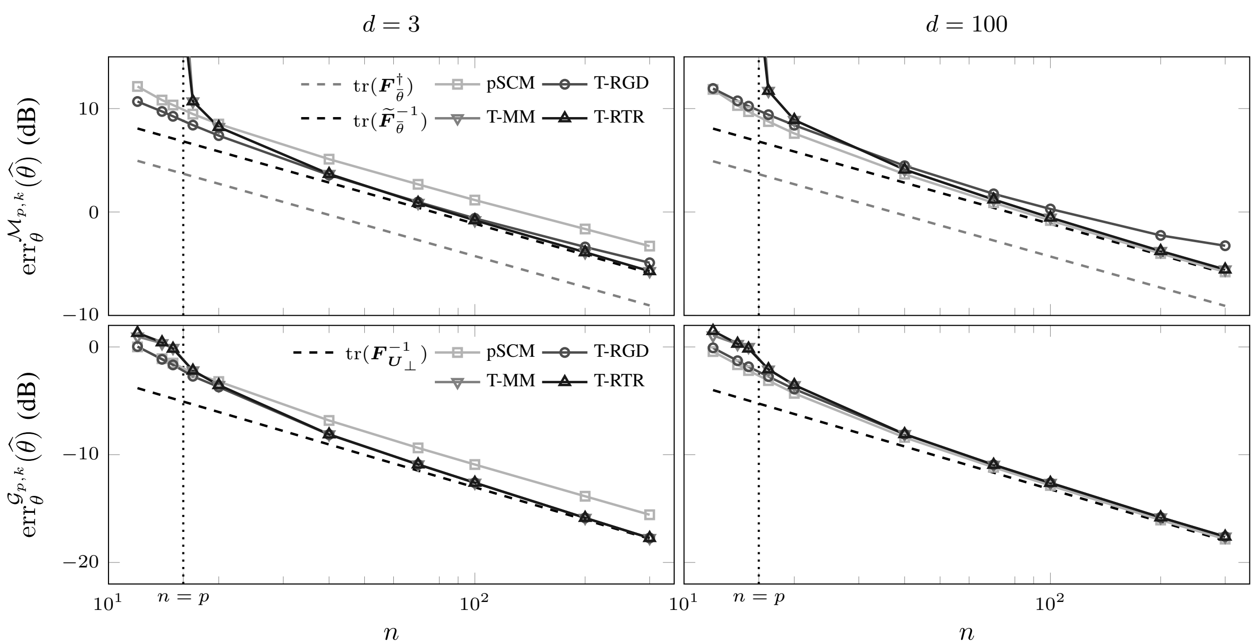


Fig. 3: Mean of error measures in propositions 10 (top) and 11 (bottom) of methods pSCM, T-MM, T-RGD and T-RTR along with corresponding intrinsic Cramér-Rao bounds in (23) and propositions 10 and 11 as functions of the number of samples $n$. Means are computed over 500 simulated sets $\{\boldsymbol{x}_i\}_{i=1}^n$ with $d = 3$ (left) and 100 (right), $p = 16$ and $k = 8$.

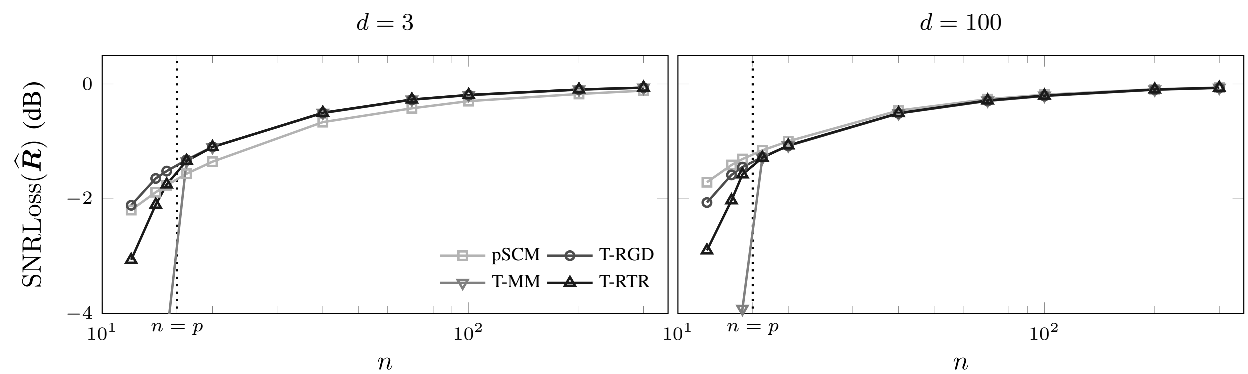


Fig. 4: SNRLoss of adaptive filters built from pSCM, T-MM, T-RGD and T-RTR as functions of the number of samples $n$. Means are computed over 500 simulated sets $\{\boldsymbol{x}_i\}_{i=1}^n$ with $d = 3$ (left) and 100 (right), $p = 16$ and $k = 8$.


[15] M. Rangaswamy, F. C. Lin, and K. R. Gerlach. Robust adaptive signal processing methods for heterogeneous radar clutter scenarios. *Signal Processing*, 84(9):1653–1665, 2004.

[16] D. Ruppert. *Statistics and data analysis for financial engineering*, volume 13. Springer, 2011.

[17] B. Kang, V. Monga, and M. Rangaswamy. Rank-constrained maximum likelihood estimation of structured covariance matrices. *IEEE Transactions on Aerospace and Electronic Systems*, 50(1):501–515, 2014.

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

[19] A. Wiesel. Geodesic convexity and covariance estimation. *IEEE Transactions on Signal Processing*, 60(12):6182–6189, 2012.

[20] A. Wiesel. Unified framework to regularized covariance estimation in scaled gaussian models. *IEEE Transactions on Signal Processing*, 60(1):29–38, 2012.

[21] E. Ollila and D. E. Tyler. Regularized $M$-estimators of scatter matrix. *IEEE Transactions on Signal Processing*, 62(22):6059–6070, 2014.

[22] L. Duembgen and D. E. Tyler. Geodesic convexity and regularized scatter estimators. *arXiv preprint arXiv:1607.05455*, 2016.

[23] S. Bonnabel and R. Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1055–1070, 2009.

[24] B. Vandereycken and S. Vandewalle. A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2553–2579, 2010.

[25] G. Meyer, S. Bonnabel, and R. Sepulchre. Regression on fixed-rank positive semidefinite matrices: a Riemannian approach. *Journal of Machine Learning Research*, 12:593–625, 2011.

[26] B. Vandereycken, P.-A. Absil, and S. Vandewalle. A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank. *IMA Journal of Numerical Analysis*, 33(2):481–514, 2012.

[27] E. Massart and P.-A. Absil. Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices. *Technical Report UCL-INMA-2018.06*, 2018.

[28] A. Edelman, T.A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[29] A. Breloy, G. Ginolhac, A. Renaux, and F. Bouchard. Intrinsic Cramér–Rao bounds for scatter and shape matrices estimation in CES distributions. *IEEE Signal Processing Letters*, 26(2):262–266, 2018.

[30] F. Pascal, P. Forster, J.P. Ovarlez, and P. Larzabal. Performance analysis of covariance matrix estimates in impulsive noise. *Signal Processing, IEEE Transactions on*, 56(6):2206–2217, June 2008.

[31] M. Mahot, F. Pascal, P. Forster, and J.-P. Ovarlez. Asymptotic properties of robust complex covariance matrix estimates. *IEEE Transactions on Signal Processing*, 61(13):3348–3356, July 2013.

[32] Y. Sun, P. Babu, and D. P. Palomar. Robust estimation of structured

covariance matrix for heavy-tailed elliptical distributions. *IEEE Transactions on Signal Processing*, 64(14):3576–3590, 2016.

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

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

[35] O. Besson and Y. I. Abramovich. On the Fisher information matrix for multivariate elliptically contoured distributions. *IEEE Signal Processing Letters*, 20(11):1130–1133, 2013.

[36] M. Greco and F. Gini. Cramér-Rao lower bounds on covariance matrix estimation for complex elliptically symmetric distributions. *IEEE Transactions on Signal Processing*, 61(24):6401–6409, 2013.

[37] F. Pascal and A. Renaux. Statistical analysis of the covariance matrix MLE in K-distributed clutter. *Signal Processing*, 90(4):1165–1175, 2010.

[38] J. D. Gorman and A. O. Hero. Lower bounds for parametric estimation with constraints. *IEEE Transactions on Information Theory*, 36(6):1285–1301, 1990.

[39] T. L. Marzetta. A simple derivation of the constrained multiple parameter Cramér-Rao bound. *IEEE Transactions on Signal Processing*, 41(6):2247–2249, 1993.

[40] P. Stoica and B. C. Ng. On the Cramér-Rao bound under parametric constraints. *IEEE Signal Processing Letters*, 5(7):177–179, 1998.

[41] N. Boumal. On intrinsic Cramér-Rao bounds for Riemannian submanifolds and quotient manifolds. *IEEE Transactions on Signal Processing*, 61(7):1809–1821, 2013.

[42] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian geometry*. Springer, 3rd edition, 2004.

[43] F. Kai-Tai and Z. Yao-Ting. *Generalized multivariate analysis*. Science Press Beijing and Springer-Verlag, Berlin, 1990.

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

[45] N. A. Goodman and J. M. Stiles. On clutter rank observed by arbitrary arrays. *IEEE Transactions on Signal Processing*, 55(1):178–186, 2007.

[46] P. Stoica and Y. Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4):36–47, 2004.

[47] R. Bhatia. *Positive definite matrices*. Princeton University Press, 2009.

[48] P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematica*, 80(2):199–220, 2004.

[49] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15:1455–1459, 2014.

[50] V. Roizman, M. Jonckheere, and F. Pascal. A flexible EM-like clustering algorithm for noisy data. *arXiv preprint arXiv:1907.01660*, 2019.

[51] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.