

Intrinsic Bayesian Cramér-Rao Bound with an Application to Covariance Matrix Estimation

Florent Bouchard, Alexandre Renaux, Guillaume Ginolhac, Arnaud Breloy

Abstract

This paper presents a new performance bound for estimation problems where the parameter to estimate lies in a Riemannian manifold (a smooth manifold endowed with a Riemannian metric) and follows a given prior distribution. In this setup, the chosen Riemannian metric induces a geometry for the parameter manifold, as well as an intrinsic notion of the estimation error measure. Performance bounds for such error measure were previously obtained in the non-Bayesian case (when the unknown parameter is assumed to deterministic), and referred to as *intrinsic* Cramér-Rao bound. The presented result then appears either as: *a*) an extension of the intrinsic Cramér-Rao bound to the Bayesian estimation framework; *b*) a generalization of the Van-Trees inequality (Bayesian Cramér-Rao bound) that accounts for the aforementioned geometric structures. In a second part, we leverage this formalism to study the problem of covariance matrix estimation when the data follow a Gaussian distribution, and whose covariance matrix is drawn from an inverse Wishart distribution. Performance bounds for this problem are obtained for both the mean squared error (Euclidean metric) and the natural Riemannian distance for Hermitian positive definite matrices (affine invariant metric). Numerical simulation illustrate that assessing the error with the affine invariant metric is revealing of interesting properties of the maximum a posteriori and minimum mean square error estimator, which are not observed when using the Euclidean metric.

Index Terms

Cramér-Rao bound, Riemannian geometry, Bayesian estimation, covariance matrix estimation

I. INTRODUCTION

Performance bounds are fundamental quantities that allow to characterize the optimal accuracy (generally assessed in terms of mean squared error) that can be achieved for a fixed setup of a given estimation

F. Bouchard and A. Renaux are with CNRS, L2S, CentraleSupélec, Université Paris-Saclay. G. Ginolhac is with LISTIC (EA3703), University Savoie Mont Blanc. A. Breloy is with CEDRIC, CNAM.

problem. Such bounds provide useful tools that can be used to evaluate the validity of estimation procedures, or to design systems so that a certain level of accuracy is actually achievable in practice. When the unknown parameter to be estimated is assumed to be deterministic, the most famous of these is the Cramér-Rao bound [1], [2], and many works addressed its generalization and refinements, e.g., to predict so-called threshold phenomena (sometimes observed at low sample support and/or low SNR when the observation model is non-linear) in estimation performance [3]–[7]. In Bayesian estimation, the unknown parameter is rather assumed to follow a known distribution, that reflects some prior knowledge. A well-known performance bound for this context is provided by the Van Trees inequality [8], also referred to as the Bayesian Cramér-Rao bound. As for the deterministic case, many variations, generalizations, and refinements were proposed in the literature [5], [9]–[11].

In practice, the parameter to be estimated can often satisfy a set of constraints dictated by the model assumptions (e.g., a fixed normalization). If these constraints can be expressed in a system of equations, the so-called constrained Cramér-Rao bound [12]–[15] generalize the standard (i.e. non-Bayesian) Cramér-Rao bound to account for this inherent structure. It was further studied in [16]–[18] (e.g., to account for estimation bias), and extended to other classes of bounds in [19], [20]. The extension to the Bayesian context was less studied in the literature. One can cite [21] which deals with discrete-time nonlinear filtering, or [22] where random constraints were addressed. Unfortunately, in many cases, the set of constraints cannot easily be expressed in a system of equations, but actually yields a smooth manifold (e.g., linear subspaces, and covariance matrices). It becomes then possible to leverage tools from Riemannian geometry, which offers several advantages to derive new performance bounds [23]–[35]. Notably, this paper will focus on the formalism of intrinsic (non-Bayesian) Cramér-Rao bounds proposed in [24] (see also introductions in [26], [29], [35]). This work proposed a generalization of the (non-Bayesian) Cramér-Rao inequality that is obtained when the error vector is defined from the Riemannian logarithm. Depending on the choice of the Riemannian metric, this results allows for bounding a geometric distance rather than the usual mean squared error. As expected, the intrinsic Cramér-Rao bound coincides with the standard Cramér-Rao bound in the Euclidean case (i.e., the parameter lies in \mathbb{R}^d endowed with the Euclidean metric). However, in the non-Euclidean geometry, it can reveal unexpected properties that are inherent to the estimation problem and its geometry (intrinsic bias, effect of the curvature of the manifold, etc.). These properties can provide a better assessment of the estimation performance actually experienced in practice, which motivates further development of this framework.

In this work, we propose to extend some of the results of [24] to a Bayesian estimation context. The contributions of this paper are the following:

- We derive a general intrinsic counterpart of the Bayesian Cramér-Rao bound, i.e., a performance

bound for estimation problems where the parameter to estimate lies in a Riemannian manifold (with any arbitrary chosen Riemannian metric) and follows an assumed prior distribution. To the best of our knowledge, few works addressed this topic: [36] derived a scalar inequality based on the Bayesian Cramér–Rao bound for smooth loss functions on manifolds, and [37] considered re-obtaining the Bayesian Cramér–Rao bound from the information theory perspective [38] i.e. the Riemannian geometry of the parameter space induced by metrics related to the statistical model (notably, the Fisher information metric is derived from the Hessian of the Kullback-Leibler divergence). Closer to the formalism considered in this paper (i.e., defining the error vector from the Riemannian logarithm), intrinsic Cramér–Rao bounds [30], variants [39], and intrinsic Bayesian Cramér–Rao bounds [33], [40] were proposed for parameter that belong to Lie groups (smooth manifolds with an additional group structure).

- We study the problem of covariance matrix estimation when the data is sampled from a Gaussian distribution, and whose covariance matrix is drawn from an inverse Wishart distribution [41]. The inverse Wishart distribution being the conjugate prior of the multivariate Gaussian distribution, it has been leveraged in numerous references from the statistics [41]–[44] and signal processing [45]–[52] literature. Thanks to the obtained intrinsic Bayesian Cramér–Rao bound, performance bounds for this problem are obtained for both the mean squared error (Euclidean metric) and the natural Riemannian distance for Hermitian positive definite matrices (affine invariant metric). Both results are new to the best of our knowledge. As in [24], the study of a proper geometry for the parameter space brings an interesting perspective to this problem. In the deterministic case, it was shown that the maximum likelihood estimator is biased and not efficient when investigating the affine invariant metric, while being apparently unbiased and efficient when evaluating error with the Euclidean metric. Conversely in the Bayesian case, we observe that the maximum a posteriori and the minimum mean square error estimators are asymptotically efficient in the intrinsic case, while these appear not to be in the Euclidean one.

The rest of the article is organized as follows: Section II contains the background related to Bayesian and intrinsic Cramér–Rao bounds, as well as necessary elements of Riemannian geometry. Section III presents the derivation of the intrinsic Bayesian Cramér–Rao bound. Section IV is dedicated to the analysis of intrinsic Bayesian Cramér–Rao bounds for the aforementioned covariance matrix estimation problem. Finally, section V shows simulations in which the obtained results allows to exhibit interesting properties of the Bayesian estimators.

Notations: *Italic type* indicates a scalar quantity, **lower case boldface** indicates a vector quantity, and

upper case boldface a matrix. The transpose conjugate operator is H and the conjugate one is $*$. $\text{Tr}(\cdot)$ and $|\cdot|$ are respectively the trace and the determinant operators. \mathcal{H}_p^{++} is the manifold of symmetric positive definite (SPD) matrices of size $p \times p$. The notation $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive definite. A complex-valued random Gaussian vector of mean $\boldsymbol{\nu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\nu}, \boldsymbol{\Sigma})$.

II. BAYESIAN AND INTRINSIC CRAMÉR-RAO LOWER BOUNDS

In this background section, we first recall the classical inequalities allowing to obtain the Cramér-Rao lower bound and its Bayesian counterpart. We then present elements of Riemannian geometry and the corresponding intrinsic Cramér-Rao lower bound, which is a generalization of the Cramér-Rao inequality obtained for a parameter lying in a Riemannian manifold [24].

A. Bayesian Cramér-Rao lower bound (Van Trees inequality)

Let $\mathbf{y} \in \mathbb{C}^p$ be an observation vector depending on an unknown parameter vector¹ $\boldsymbol{\theta} \in \mathbb{R}^d$ through its likelihood function denoted $f(\mathbf{y}|\boldsymbol{\theta})$. When the unknown parameter vector $\boldsymbol{\theta}$ is assumed to be deterministic, the Cramér-Rao lower bound [1], [2] states that any unbiased estimator $\hat{\boldsymbol{\theta}}$ satisfies

$$\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \succeq \mathbf{F}_{\boldsymbol{\theta}}^{-1}, \quad (1)$$

where

$$\mathbf{F}_{\boldsymbol{\theta}} = \mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}} \left[\frac{\partial \log f(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log f(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right] \quad (2)$$

is the Fisher information matrix (assumed to exist and to be non-singular throughout this paper). This inequality can be reinterpreted in terms of Euclidean distance between the estimator and the parameter, as

$$\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}}[d_E^2(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})] \geq \text{Tr}(\mathbf{F}_{\boldsymbol{\theta}}^{-1}), \quad (3)$$

where $d_E(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|$ is the standard Euclidean distance.

In a Bayesian setting, the parameter vector is assumed to follow a prior distribution denoted $f_{\text{prior}}(\boldsymbol{\theta})$ and the joint distribution of the couple $\{\mathbf{y}, \boldsymbol{\theta}\}$ is denoted $f(\mathbf{y}, \boldsymbol{\theta}) = f(\mathbf{y}|\boldsymbol{\theta})f_{\text{prior}}(\boldsymbol{\theta})$. The score function is defined as

$$s(\mathbf{y}, \boldsymbol{\theta}) = \frac{\partial \log f(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad (4)$$

¹Notice that if the parameter vector has several complex-valued entries, it is always possible to split their estimation problem in terms of real and imaginary parts, so we can assume without loss of generality that $\boldsymbol{\theta}$ is a real vector (with properly adjusted dimension d).

and the corresponding Bayesian Fisher information matrix is defined as

$$\mathbf{F}_B = \mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}} [s(\mathbf{y}, \boldsymbol{\theta}) s(\mathbf{y}, \boldsymbol{\theta})^T]. \quad (5)$$

Since the joint log-likelihood can be expressed $\log f(\mathbf{y}, \boldsymbol{\theta}) = \log f(\mathbf{y}|\boldsymbol{\theta}) + \log f_{\text{prior}}(\boldsymbol{\theta})$, we can split the score function (4) into two components. This results in a decomposition of the Bayesian Fisher information matrix as

$$\mathbf{F}_B = \mathbb{E}_{\boldsymbol{\theta}} [\mathbf{F}_{\boldsymbol{\theta}}] + \mathbf{F}_{\text{prior}}, \quad (6)$$

where

$$\mathbf{F}_{\text{prior}} = \mathbb{E}_{\boldsymbol{\theta}} \left[\frac{\partial \log f_{\text{prior}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log f_{\text{prior}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right]. \quad (7)$$

From these definitions, a Bayesian counterpart of the Cramér-Rao lower bound, referred to as the Van Trees inequality [8], [9], is stated in the following theorem.

Theorem 1 (Van Trees inequality [8], [9]). *Let $\hat{\boldsymbol{\theta}}$ be an estimator of $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_i$ (resp. $\hat{\boldsymbol{\theta}}_i$) denote the i^{th} element of $\boldsymbol{\theta}$ (resp. $\hat{\boldsymbol{\theta}}$). From [10], if $f(\mathbf{y}, \boldsymbol{\theta})$ is absolutely continuous with respect to $\boldsymbol{\theta}$ a.e. \mathbf{y} , if \mathbf{F}_B in (7) exists and is non-singular, and if the assumption*

$$\lim_{\boldsymbol{\theta}_i \rightarrow \pm\infty} \boldsymbol{\theta}_i f(\mathbf{y}, \boldsymbol{\theta}) = 0 \quad (8)$$

holds for all element $i \in \llbracket 1, d \rrbracket$ and for any (a.e.) $\mathbf{y} \in \mathbb{C}^p$, then

$$\mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}} [(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \succeq \mathbf{F}_B^{-1}, \quad (9)$$

which translates in expected Euclidean distance as

$$\mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}} [d_E(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})^2] \geq \text{Tr} \left((\mathbb{E}_{\boldsymbol{\theta}} [\mathbf{F}_{\boldsymbol{\theta}}] + \mathbf{F}_{\text{prior}})^{-1} \right). \quad (10)$$

B. Intrinsic Cramér-Rao bound

Intrinsic Cramér-Rao lower bound refers to the seminal work of [24], that derived a generalization of the Cramér-Rao inequality (1) in the case where $\boldsymbol{\theta}$ is assumed to be deterministic and to lie in a Riemannian manifold. To introduce this result, we first present some notations and tools of Riemannian geometry. The inequality of [24] then lies in a careful transposition of the notion of estimation error vector to this context. For more detailed coverage of differential geometry, one can refer to the standard textbooks on the topic [53]–[55]. The notations in this article are mostly inspired from the books [56], [57], which provide very good (optimization-oriented) entry points to smooth manifolds and Riemannian geometry. Lastly, a more detailed introduction to information geometry and intrinsic Cramér-Rao bound is proposed in [35].

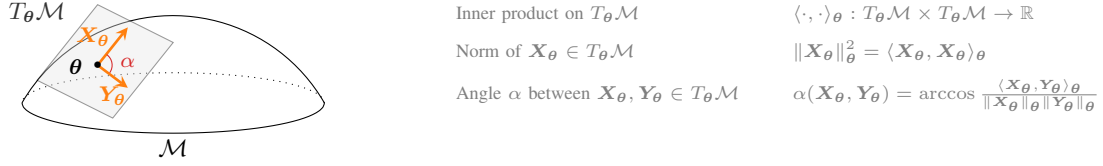


Fig. 1: A visual representation of a smooth manifold \mathcal{M} , its tangent space $T_\theta \mathcal{M}$ at point θ , and two tangent vectors \mathbf{X}_θ and \mathbf{Y}_θ . The metric $\langle \cdot, \cdot \rangle_\theta$ is an inner product on $T_\theta \mathcal{M}$ and it induces the notion of length and angle.

1) *Riemannian geometry*: A smooth manifold \mathcal{M} is a space that is locally diffeomorphic to a vector space and that possesses a differential structure. Hence, each point $\theta \in \mathcal{M}$ admits a tangent space $T_\theta \mathcal{M}$. Tangent vectors in $T_\theta \mathcal{M}$ generalize directional derivatives at θ on \mathcal{M} . To turn \mathcal{M} into a Riemannian manifold, it needs to be equipped with a Riemannian metric $\langle \cdot, \cdot \rangle_\theta$, which defines smoothly varying inner products on tangent spaces. It is important to notice that the definition of the metric is a choice, that will induce – along with the geometrical constraints – a corresponding geometry for the manifold \mathcal{M} . Figure 1 illustrates the aforementioned spaces and objects.

a) *Second-order differentiation on a Riemannian manifold*: Vector fields are functions associating one tangent vector to each point on \mathcal{M} . The directional derivatives of vector fields (i.e., the generalization of the standard directional derivatives) is defined through the Levi-Civita connection $\nabla \cdot$. Such generalization is indeed needed because the tangent space changes when one moves from one point to another on a manifold. Hence, the Levi-Civita connection provides a way of differentiating that properly reflects the structure of the manifold – i.e., the effect of the geometrical constraints and chosen metric – as illustrated in Figure 2. In practice, the Levi-Civita connection is obtained thanks to the Koszul formula: given vector fields \mathbf{X}_θ , \mathbf{Y}_θ and \mathbf{Z}_θ at $\theta \in \mathcal{M}$, it is

$$2\langle \nabla_{\mathbf{Y}_\theta} \mathbf{X}_\theta, \mathbf{Z}_\theta \rangle_\theta = D\langle \mathbf{Y}_\theta, \mathbf{Z}_\theta \rangle_\theta[\mathbf{X}_\theta] + D\langle \mathbf{X}_\theta, \mathbf{Z}_\theta \rangle_\theta[\mathbf{Y}_\theta] - D\langle \mathbf{X}_\theta, \mathbf{Y}_\theta \rangle_\theta[\mathbf{Z}_\theta] \\ + \langle \mathbf{Z}_\theta, [\mathbf{X}_\theta, \mathbf{Y}_\theta] \rangle_\theta + \langle \mathbf{Y}_\theta, [\mathbf{Z}_\theta, \mathbf{X}_\theta] \rangle_\theta - \langle \mathbf{X}_\theta, [\mathbf{Y}_\theta, \mathbf{Z}_\theta] \rangle_\theta,$$

where $D[\cdot]$ denotes the directional derivative and $[\cdot, \cdot]$ the Lie bracket.

b) *Riemannian curvature*: The Levi-Civita connection allows to define the corresponding Riemannian curvature tensor, which is

$$\mathcal{R}(\mathbf{X}_\theta, \mathbf{Y}_\theta)\mathbf{Z}_\theta = \nabla_{\mathbf{X}_\theta} \nabla_{\mathbf{Y}_\theta} \mathbf{Z}_\theta - \nabla_{\mathbf{Y}_\theta} \nabla_{\mathbf{X}_\theta} \mathbf{Z}_\theta - \nabla_{[\mathbf{X}_\theta, \mathbf{Y}_\theta]} \mathbf{Z}_\theta. \quad (11)$$

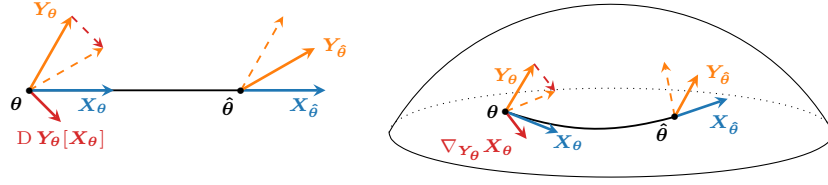


Fig. 2: Illustration of directional derivative $D Y_\theta[X_\theta]$ (left) and Levi-Civita connection $\nabla_{Y_\theta} X_\theta$ (right) of a vector field Y_θ in the direction X_θ at θ . As the directional derivative, an Levi-Civita connection describes how the vector field Y_θ evolves in a given direction X_θ . In addition, the affine connection takes into account the structure of the manifold (curvature, and non-constant metric).

It measures how second-order differentiation fails to commute on the Riemannian manifold \mathcal{M} . Indeed, it is the difference between $\nabla_{X_\theta} \nabla_{Y_\theta} \cdot - \nabla_{Y_\theta} \nabla_{X_\theta} \cdot$ and $\nabla_{[X_\theta, Y_\theta]} \cdot$. As for the Levi-Civita connection, it reflects the effects of both the geometrical constraints and the (non-constant) Riemannian metric.

c) Geodesics and Riemannian distance: The Levi-Civita connection also yields geodesics on \mathcal{M} , which generalize the concept of straight lines. The geodesic $\gamma : [0, 1] \rightarrow \mathcal{M}$ such that $\gamma(0) = \theta$ and $\dot{\gamma}(0) = X_\theta$ is defined through the differential equation

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = \mathbf{0}, \quad (12)$$

and is illustrated in Figure 3. From there, one can define the Riemannian exponential mapping. At θ , it is the mapping $\exp_\theta : T_\theta \mathcal{M} \rightarrow \mathcal{M}$ such that, for all $X_\theta \in T_\theta \mathcal{M}$, $\exp_\theta(X_\theta) = \gamma(1)$. Its inverse, the Riemannian logarithm is, at $\theta \in \mathcal{M}$, the mapping $\log_\theta : \mathcal{M} \rightarrow T_\theta \mathcal{M}$ such that, for $\hat{\theta} \in \mathcal{M}$, $\log_\theta(\hat{\theta}) = X_\theta$ where $\exp_\theta(X_\theta) = \hat{\theta}$. The Riemannian exponential and logarithm mappings are illustrated in Figure 3. The manifold \mathcal{M} is said to be complete if for all $\theta \in \mathcal{M}$ the Riemannian exponential is well defined on the entire tangent space $T_\theta \mathcal{M}$. Otherwise we can restrict this operator to be defined at each point only for tangent vectors whose norm are lower than the so-called injectivity radius (largest radius for which the exponential map at θ is a diffeomorphism). In any case, the Riemannian logarithm \log_θ might not be uniquely defined (e.g., multiple windings for great-circle geodesics on the sphere). In some cases, one can obtain a unique definition for this operator by choosing the tangent vector of shortest length. In this study, we assume that \exp_θ and \log_θ are well defined on the manifold. As explained in [24], this can be extended to the case of conjugate points for which \exp_θ is singular because such points have measure zero. Finally, the Riemannian distance $d_R : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ is

$$d_R(\theta, \hat{\theta}) = \int_0^1 \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}} dt, \quad (13)$$

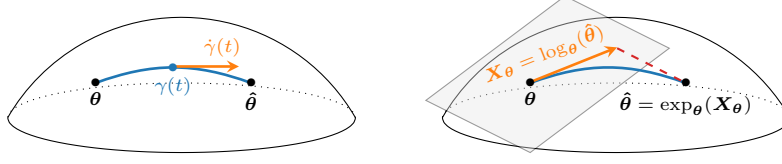


Fig. 3: Illustration of geodesics (left), Riemannian exponential and logarithm mappings (right). The Riemannian distance $d_R(\theta, \hat{\theta})$ is the length of the geodesic joining θ and $\hat{\theta}$.

where γ is the geodesic on \mathcal{M} such that $\gamma(0) = \theta$ and $\gamma(1) = \hat{\theta}$.

2) *From Riemannian geometry to performance bound:* We consider the problem of estimating a parameter θ in a manifold \mathcal{M} from some data \mathbf{y} in \mathbb{C}^p . In the considered setting, the data at hand are drawn from a distribution with probability density function $f(\mathbf{y}|\theta)$. As for the classical Euclidean case, one wishes to establish a lower performance bound for the estimation problem. This is achieved by the so-called intrinsic Cramér-Rao lower bound [24], [29].

a) *Estimation error:* First, one needs to re-define the estimation error of an unbiased estimator $\hat{\theta} \in \mathcal{M}$ of the true parameter $\theta \in \mathcal{M}$. Indeed, the notion of subtraction “ $\hat{\theta} - \theta$ ” from the Euclidean case described in Section II is not defined intrinsically on \mathcal{M} . A proper definition of the error that is in accordance with the manifold \mathcal{M} and the geometry induced by the chosen Riemannian metric $\langle \cdot, \cdot \rangle$, is provided by the Riemannian logarithm: an element of the tangent space $T_\theta \mathcal{M}$ of θ that “points towards” $\hat{\theta}$, and whose norm corresponds to the Riemannian distance $d_R(\theta, \hat{\theta})$ (cf. Section and Section II-B1c and Figure 3). Furthermore, it will be helpful to handle this object using a system of coordinates [30]. To do so, let $\{\Omega_i\}_{i=1}^d$ be an orthonormal basis of $T_\theta \mathcal{M}$, where d is the dimension of \mathcal{M} . In practice, such a basis can be obtained either analytically from mathematical calculations or numerically, thanks to the Gram-Schmidt orthonormalization process. This basis yields the decomposition

$$\log_\theta(\hat{\theta}) = \sum_{i=1}^d (\varepsilon_\theta)_i \Omega_i. \quad (14)$$

The vector $\varepsilon_\theta = [(\varepsilon_\theta)_i]_{i=1}^d \in \mathbb{R}^d$ is the coordinates error vector, which is obtained as

$$(\varepsilon_\theta)_i = \langle \log_\theta(\hat{\theta}), \Omega_i \rangle_\theta. \quad (15)$$

The covariance matrix of this error measure is $C_\theta = \mathbb{E}_\mathbf{y}[\varepsilon_\theta \varepsilon_\theta^T]$. Further note that the squared norm of this vector ε_θ – which is the trace of C_θ – corresponds to the squared Riemannian distance between θ and $\hat{\theta}$, i.e.,

$$d_R^2(\theta, \hat{\theta}) = \langle \log_\theta(\hat{\theta}), \log_\theta(\hat{\theta}) \rangle_\theta = \|\varepsilon_\theta\|_2^2 = \text{Tr}(C_\theta). \quad (16)$$

b) *Fisher information matrix*: To obtain an inequality such as (1), we need to redefine the Fisher information matrix as follows: it is the matrix $\mathbf{F}_\theta \in \mathbb{R}^{d \times d}$ whose ij^{th} element is

$$(\mathbf{F}_\theta)_{ij} = \langle \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j \rangle_\theta^{\text{FIM}}, \quad (17)$$

where $\langle \cdot, \cdot \rangle_\theta^{\text{FIM}}$ is the Fisher information metric. As explained in [24], this metric is defined, for all $\theta \in \mathcal{M}$, $\mathbf{X}_\theta, \mathbf{Y}_\theta \in T_\theta \mathcal{M}$ as

$$\begin{aligned} \langle \mathbf{X}_\theta, \mathbf{Y}_\theta \rangle_\theta^{\text{FIM}} &= \mathbb{E}_{\mathbf{y}}[\mathbf{D}_\theta \log f(\mathbf{y}|\theta)[\mathbf{X}_\theta] \cdot \mathbf{D}_\theta \log f(\mathbf{y}|\theta)[\mathbf{Y}_\theta]] \\ &= -\mathbb{E}_{\mathbf{y}}[\mathbf{D}_\theta^2 \log f(\mathbf{y}|\theta)[\mathbf{X}_\theta, \mathbf{Y}_\theta]], \end{aligned} \quad (18)$$

where $\mathbf{D}^2 \cdot [\cdot, \cdot]$ is the second order directional derivative².

Remark 1. Notice that we consider a geometry and notion of estimation error induced by any chosen metric $\langle \cdot, \cdot \rangle$. In particular, the geometry induced by the Fisher information metric $\langle \cdot, \cdot \rangle_\theta^{\text{FIM}}$ is referred to as the Fisher-Rao information geometry (of the considered statistical model) [35]. Interestingly, the choice $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_\theta^{\text{FIM}}$ always yields $\mathbf{F}_\theta = \mathbf{I}_d$, which underlines the inherent match between the Fisher information metric with the estimation problem. However, the Fisher-Rao information geometry might not always be tractable, or might not always be of interest for measuring the estimation performance in some contexts. Hence, being able to rely on any arbitrary metric $\langle \cdot, \cdot \rangle$ is a key point in this work.

c) *Intrinsic Cramér-Rao lower bound*: The Fisher information matrix \mathbf{F}_θ can be leveraged to define a performance lower bound on the covariance \mathbf{C}_θ of the coordinates error vector $\boldsymbol{\varepsilon}_\theta$, it yields the inequality stated in Theorem 2.

Theorem 2 (intrinsic Cramér-Rao lower bound [24]). *Let $\hat{\theta}$ be an unbiased estimator of θ in \mathcal{M} . The intrinsic Cramér-Rao lower bound of the covariance $\mathbf{C}_\theta = \mathbb{E}_{\mathbf{y}}[\boldsymbol{\varepsilon}_\theta \boldsymbol{\varepsilon}_\theta^T]$ of the error vector defined in (15) is*

$$\mathbf{C}_\theta \succeq \mathbf{F}_\theta^{-1} - \frac{1}{3}(\mathbf{F}_\theta^{-1} \mathcal{R}_m(\mathbf{F}_\theta) + \mathcal{R}_m(\mathbf{F}_\theta) \mathbf{F}_\theta^{-1}), \quad (19)$$

where the Fisher matrix \mathbf{F}_θ is defined in (17) and \mathcal{R}_m is a curvature related term such that, for $i, j \in \{1, \dots, d\}$,

$$(\mathcal{R}_m(\mathbf{C}_\theta))_{i,j} = \mathbb{E}_{\mathbf{y}}[\langle \mathcal{R}(\log_\theta(\hat{\theta}), \boldsymbol{\Omega}_i) \boldsymbol{\Omega}_j, \log_\theta(\hat{\theta}) \rangle_\theta]. \quad (20)$$

Taking the trace of this inequality further yields

$$\mathbb{E}_{\mathbf{y}}[d_R^2(\hat{\theta}, \theta)] \geq \text{Tr}(\mathbf{F}_\theta^{-1} - \frac{1}{3}(\mathbf{F}_\theta^{-1} \mathcal{R}_m(\mathbf{F}_\theta) + \mathcal{R}_m(\mathbf{F}_\theta) \mathbf{F}_\theta^{-1})). \quad (21)$$

²When differentiating functions with several variables such as $f(\mathbf{y}|\theta)$, we indicate the variable that is differentiated with a subscript, i.e., $\mathbf{D}_\theta \log f(\mathbf{y}|\theta)[\cdot]$ is the directional derivative of $\log f(\mathbf{y}|\theta)$ with respect to the variable θ .

Remark 2 (From intrinsic (Riemannian) to Euclidean Cramér-Rao bound). *If we consider the manifold \mathcal{M} as the Euclidean space \mathbb{R}^d equipped with the Euclidean metric, then the Cramér-Rao inequality (19) from Theorem 2 is equal to (1). Indeed, in this case, the logarithm mapping is simply $\log_{\theta}(\hat{\theta}) = \hat{\theta} - \theta$, hence the error vector is $\varepsilon_{\theta} = \hat{\theta} - \theta$. The Fisher information matrix (17) is then reduced to definition (2). Finally, \mathbb{R}^d is flat and the curvature tensor (11) is equal to zero.*

III. INTRINSIC BAYESIAN CRAMÉR-RAO BOUND

In this section, an intrinsic Bayesian Cramér-Rao lower bound is proposed. More specifically, an intrinsic derivation of the Van-Trees inequality is presented. It consists in the generalization of Theorem 1 for an error vector defined from objects on a manifold \mathcal{M} as in Section II-B. This section is divided in two subsections. Section III-A contains the required assumptions. Section III-B presents the intrinsic Bayesian Cramér-Rao lower bound along with its proof.

A. Hypotheses

To derive the intrinsic Bayesian Cramér-Rao bound stated in Theorem 3, we first need several assumptions on the model. These ensure that the computed derivatives, expectations, and matrix inversions are indeed well defined. These assumptions are provided below, along with some explanations and interpretations.

Assumption 1. *$f(\mathbf{y}, \theta)$ is absolutely continuous with respect to θ for any \mathbf{y} .*

Assumptions 1 simply state a standard regularity condition that is used when studying Bayesian bounds (cf. [10]).

Assumption 2. *The parameterization of chosen orthonormal bases $\{\Omega_i\}_{i=1}^d$ of tangent spaces $T_{\theta}\mathcal{M}$ is smooth with respect to θ .*

Assumption 2 is simply technical and usually satisfied: bases are in practice constructed with canonical elements and smooth functions of θ to ensure orthonormality with respect to the chosen metric $\langle \cdot, \cdot \rangle$.

Assumption 3. *The manifold \mathcal{M} is measurable, and the function $\mu(\cdot)$ defines a measure for this space.*

Assumption 3 allows to define the expectation over θ as

$$\mathbb{E}_{\theta}[g(\theta)] = \int_{\theta \in \mathcal{M}} g(\theta) f_{\text{prior}}(\theta) \mu(d\theta). \quad (22)$$

Notice that, in the general case, the notion of integration over a manifold is not trivial, and conditions the design of f_{prior} , that needs to integrate to one over \mathcal{M} . In this setup, even computing the normalizing

constant is often challenging. Examples of such distributions include Wishart-based priors (cf. Section IV), distributions on orthonormal frames and subspaces (cf. [58] and references therein), Riemannian Gaussian distribution [59], and concentrated Gaussian distributions on Lie groups [39].

Assumption 4. F_B defined for each element by $(F_B)_{i,j} = \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[D_{\boldsymbol{\theta}} \log f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_i] \cdot D_{\boldsymbol{\theta}} \log f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_j]]$ exists and is non-singular.

As for assumption 1, assumption 4 states a standard regularity condition that is used when studying Bayesian bounds (cf. [10]).

Assumption 5. For all $i, j \in \{1, \dots, d\}$, the relationship

$$\int_{\mathcal{M}} \int_{\mathbb{C}^p} \langle \nabla_{\boldsymbol{\Omega}_j} (f(\mathbf{y}, \boldsymbol{\theta}) \log_{\boldsymbol{\theta}} \hat{\boldsymbol{\theta}}), \boldsymbol{\Omega}_i \rangle_{\boldsymbol{\theta}} d\mathbf{y} \mu(d\boldsymbol{\theta}) = 0 \quad (23)$$

is satisfied.

Assumption 5 also appears as rather technical, but is related to assumption (8) in the usual derivation of the Bayesian Cramér-Rao bound. Indeed a proof strategy of Theorem 1 involves an integration by parts, requiring to state

$$\int_{\mathbb{R}^d} \int_{\mathbb{C}^p} \frac{\partial(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) f(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} d\mathbf{y} d\boldsymbol{\theta} = \mathbf{0} \quad (24)$$

before concluding³. Assumption 5 is the Riemannian counterpart of (24), i.e., it states the same condition when using the formalism of intrinsic Cramér-Rao bound detailed in Section III. Further notice that in the case where \mathcal{M} is a unimodular Lie group [39], an equivalent condition to Assumption 5 is required and always satisfied – see Equation (11) in [39].

B. Main theorem

We are now able to provide our main result, i.e., an intrinsic derivation of the Van Trees inequality, in Theorem 3. As for the usual intrinsic Cramér-Rao bound, it extends the (Euclidean) Bayesian bound from Theorem 1 to the context of a Riemannian manifold \mathcal{M} .

Theorem 3 (Intrinsic Bayesian Cramér-Rao lower bound). *Let $\hat{\boldsymbol{\theta}}$ be an estimator of a parameter $\boldsymbol{\theta} \in \mathcal{M}$ constructed from $\mathbf{y} \in \mathbb{C}^p$, with $\mathbf{y} \simeq f(\mathbf{y}|\boldsymbol{\theta})$ and $\boldsymbol{\theta} \sim f_{\text{prior}}(\boldsymbol{\theta})$. Under assumptions 1-5, the covariance matrix of the error $\mathbf{C} = \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\boldsymbol{\varepsilon}_{\boldsymbol{\theta}} \boldsymbol{\varepsilon}_{\boldsymbol{\theta}}^T]$ satisfies the following inequality*

$$\mathbf{C} + \frac{1}{3}(\mathbf{F}_B^{-1} R_m(\mathbf{C}) + R_m(\mathbf{C}) \mathbf{F}_B^{-1}) \succeq \mathbf{F}_B^{-1}, \quad (25)$$

³In [10], the authors used assumption (8) only since their work is conducted for a special case of a parameter vector lying on \mathbb{R}^d .

in which the Bayesian Fisher matrix \mathbf{F}_B is defined with

$$\begin{aligned} (\mathbf{F}_B)_{i,j} &= \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\mathbf{D}_{\boldsymbol{\theta}} \log f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_i] \cdot \mathbf{D}_{\boldsymbol{\theta}} \log f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_j]] \\ &= \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[-\mathbf{D}_{\boldsymbol{\theta}}^2 \log f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j]], \end{aligned} \quad (26)$$

and where the curvature-related term R_m defined from \mathbf{C} as

$$(R_m(\mathbf{C}))_{i,j} = \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\langle \mathcal{R}(\log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Omega}_i) \boldsymbol{\Omega}_j, \log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \rangle_{\boldsymbol{\theta}}]. \quad (27)$$

Proof. As in the Euclidean case, we start by defining the score $\mathbf{s}_{\boldsymbol{\theta}} \in \mathbb{R}^d$, whose i^{th} element is defined as

$$(\mathbf{s}_{\boldsymbol{\theta}})_i = \mathbf{D}_{\boldsymbol{\theta}} \log f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_i]. \quad (28)$$

We also define the Fisher matrix as $\mathbf{F}_B = \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\mathbf{s}_{\boldsymbol{\theta}} \mathbf{s}_{\boldsymbol{\theta}}^T]$ – also defined in (26). As for the Euclidean derivation of the Bayesian bound, we set

$$\mathbf{v}_{\boldsymbol{\theta}} = \boldsymbol{\varepsilon}_{\boldsymbol{\theta}} - \mathbf{F}_B^{-1} \mathbf{s}_{\boldsymbol{\theta}}, \quad (29)$$

where $\boldsymbol{\varepsilon}_{\boldsymbol{\theta}}$ is the coordinates error vector defined in (15). By definition, we have $\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\mathbf{v}_{\boldsymbol{\theta}} \mathbf{v}_{\boldsymbol{\theta}}^T] \succeq \mathbf{0}$. Hence,

$$\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\mathbf{v}_{\boldsymbol{\theta}} \mathbf{v}_{\boldsymbol{\theta}}^T] = \mathbf{C} + \mathbf{F}_B^{-1} - \mathbf{F}_B^{-1} \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\mathbf{s}_{\boldsymbol{\theta}} \boldsymbol{\varepsilon}_{\boldsymbol{\theta}}^T] - \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\boldsymbol{\varepsilon}_{\boldsymbol{\theta}} \mathbf{s}_{\boldsymbol{\theta}}^T] \mathbf{F}_B^{-1} \succeq \mathbf{0}, \quad (30)$$

where $\mathbf{C} = \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\boldsymbol{\varepsilon}_{\boldsymbol{\theta}} \boldsymbol{\varepsilon}_{\boldsymbol{\theta}}^T]$. As in the Euclidean case, the key issue is the computation of $\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\mathbf{s}_{\boldsymbol{\theta}} \boldsymbol{\varepsilon}_{\boldsymbol{\theta}}^T]$ (or equivalently $\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\boldsymbol{\varepsilon}_{\boldsymbol{\theta}} \mathbf{s}_{\boldsymbol{\theta}}^T]$). Let $\mathbf{A} = \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\boldsymbol{\varepsilon}_{\boldsymbol{\theta}} \mathbf{s}_{\boldsymbol{\theta}}^T]$. Each element $(i, j) \in \{1, \dots, d\}^2$ is defined as

$$\mathbf{A}_{i,j} = \int_{\mathcal{M}} \int_{\mathbb{C}^p} \langle \log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Omega}_i \rangle_{\boldsymbol{\theta}} \mathbf{D}_{\boldsymbol{\theta}} \log f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_j] f(\mathbf{y}, \boldsymbol{\theta}) d\mathbf{y} \mu(d\boldsymbol{\theta}). \quad (31)$$

Since

$$\mathbf{D}_{\boldsymbol{\theta}} \log f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_j] = \frac{1}{f(\mathbf{y}, \boldsymbol{\theta})} \mathbf{D}_{\boldsymbol{\theta}} f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_j], \quad (32)$$

we get

$$\begin{aligned} \mathbf{A}_{i,j} &= \int_{\mathcal{M}} \int_{\mathbb{C}^p} \langle \log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Omega}_i \rangle_{\boldsymbol{\theta}} \mathbf{D}_{\boldsymbol{\theta}} f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_j] d\mathbf{y} \mu(d\boldsymbol{\theta}) \\ &= \int_{\mathcal{M}} \int_{\mathbb{C}^p} \langle \mathbf{D}_{\boldsymbol{\theta}} f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_j] \log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Omega}_i \rangle_{\boldsymbol{\theta}} d\mathbf{y} \mu(d\boldsymbol{\theta}). \end{aligned} \quad (33)$$

Since the Levi-Civita connection satisfies Leibniz rule, we have

$$\begin{aligned} \langle \nabla_{\boldsymbol{\Omega}_j} (f(\mathbf{y}, \boldsymbol{\theta}) \log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})), \boldsymbol{\Omega}_i \rangle_{\boldsymbol{\theta}} &= \langle \mathbf{D}_{\boldsymbol{\theta}} f(\mathbf{y}, \boldsymbol{\theta})[\boldsymbol{\Omega}_j] \log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Omega}_i \rangle_{\boldsymbol{\theta}} \\ &+ \langle f(\mathbf{y}, \boldsymbol{\theta}) \nabla_{\boldsymbol{\Omega}_j} \log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Omega}_i \rangle_{\boldsymbol{\theta}}. \end{aligned} \quad (34)$$

From there, we obtain

$$\begin{aligned} \mathbf{A}_{i,j} &= \int_{\mathcal{M}} \int_{\mathbb{C}^p} \langle \nabla_{\boldsymbol{\Omega}_j} (f(\mathbf{y}, \boldsymbol{\theta}) \log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})), \boldsymbol{\Omega}_i \rangle_{\boldsymbol{\theta}} d\mathbf{y} \mu(d\boldsymbol{\theta}) \\ &- \int_{\mathcal{M}} \int_{\mathbb{C}^p} \langle f(\mathbf{y}, \boldsymbol{\theta}) \nabla_{\boldsymbol{\Omega}_j} \log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Omega}_i \rangle_{\boldsymbol{\theta}} d\mathbf{y} \mu(d\boldsymbol{\theta}). \end{aligned} \quad (35)$$

Assumption 5 further yields

$$\begin{aligned} \mathbf{A}_{i,j} &= - \int_{\mathcal{M}} \int_{\mathbb{C}^p} \langle f(\mathbf{y}, \boldsymbol{\theta}) \nabla_{\boldsymbol{\Omega}_j} \log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Omega}_i \rangle_{\boldsymbol{\theta}} d\mathbf{y} \mu(d\boldsymbol{\theta}) \\ &= \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[\langle -\nabla_{\boldsymbol{\Omega}_j} \log_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Omega}_i \rangle_{\boldsymbol{\theta}}]. \end{aligned} \quad (36)$$

From there, the end of the proof is very similar to the ones of the (non-Bayesian) intrinsic Cramér-Rao bound that can be found in [24], [27]. From Equations (38) and (39) in [24], we have

$$\begin{aligned} \langle -\nabla_{\Omega_j} \log_{\theta}(\hat{\theta}), \Omega_i \rangle_{\theta} &= \langle \Omega_i, \Omega_j \rangle_{\theta} \\ &- \frac{1}{3} \langle \mathcal{R}(\log_{\theta}(\hat{\theta}), \Omega_j) \Omega_i, \log_{\theta}(\hat{\theta}) \rangle_{\theta} + \mathcal{O}(\|\log_{\theta}(\hat{\theta})\|^3). \end{aligned} \quad (37)$$

Recall that $\{\Omega_i\}_{i=1}^d$ is an orthonormal basis of $T_{\theta}\mathcal{M}$ (i.e., $\langle \Omega_i, \Omega_j \rangle_{\theta} = \delta_{i,j}$) and let the operator R_m defined as in (27). From combining (36) and (37) and by neglecting the higher order terms $\mathcal{O}(\|\log_{\theta}(\hat{\theta})\|^3)$, one gets

$$\mathbf{A} = \mathbb{E}_{\mathbf{y}, \theta}[\varepsilon_{\theta} \mathbf{s}_{\theta}^T] = \mathbf{I}_d - \frac{1}{3} R_m(\mathbf{C}). \quad (38)$$

Injecting this in (30), one obtains

$$\mathbf{C} - \mathbf{F}_B^{-1} + \frac{1}{3}(\mathbf{F}_B^{-1} R_m(\mathbf{C}) + R_m(\mathbf{C}) \mathbf{F}_B^{-1}) \succeq \mathbf{0}. \quad (39)$$

which concludes the proof. \square

Notice that Theorem 3 provides an inequality relating the covariance matrix of the error and the Bayesian Fisher information matrix, as in the classical case of Theorem 1. However, as for the (non-Bayesian) intrinsic Cramer-Rao bound, the result is obtained according to a chosen error metric, which brings some notable differences regarding the definition of these objects, and the involvement of curvature-related terms within the inequality. Still, Theorem 3 does not provide a practical way for bounding the expected estimation error because of the terms in $R_m(\mathbf{C})$. For small errors, the inequality can be simplified to a tractable expression in the following corollary.

Corollary 1 (First approximation of intrinsic Bayesian Cramér-Rao lower bound). *For sufficiently small errors, the inequality in Theorem 3 can be simplified to*

$$\mathbf{C} \succeq \mathbf{F}_B^{-1} - \frac{1}{3}(\mathbf{F}_B^{-1} R_m(\mathbf{F}_B^{-1}) + R_m(\mathbf{F}_B^{-1}) \mathbf{F}_B^{-1}) + \mathcal{O}(\lambda_{\max}(\mathbf{F}_B^{-1})^3) \quad (40)$$

where $\lambda_{\max}(\mathbf{F}_B^{-1})$ is the largest eigenvalue of \mathbf{F}_B^{-1} , and where the linear operator $R_m(\mathbf{M})$ is defined for any matrix $\mathbf{M} \in \mathbb{C}^{d \times d}$ [60] as

$$(R_m(\mathbf{M}))_{i,j} = \sum_{k,l} \langle \mathcal{R}(\Omega_k, \Omega_i) \Omega_j, \Omega_l \rangle_{\theta} \mathbf{M}_{k,l}. \quad (41)$$

Proof. Let the operator Δ be defined as

$$\Delta(\mathbf{C}) = \frac{1}{3}(\mathbf{F}_B^{-1} R_m(\mathbf{C}) + R_m(\mathbf{C}) \mathbf{F}_B^{-1}), \quad (42)$$

and Id denotes the identity operator. The result in Theorem 3 then reads $(\text{Id} + \Delta)(\mathbf{C}) \succeq \mathbf{F}_B^{-1}$. For small expected errors (referred to as large signal to noise ratio in [24]), \mathbf{F}_B^{-1} is negligible compared to \mathbf{I}_d , so

the operator Δ can be assumed to be a small perturbation of Id . Thus, $\text{Id} + \Delta$ is positive definite and its inverse has the first order Taylor expansion $(\text{Id} + \Delta)^{-1} = \text{Id} - \Delta + \Delta^2 - \dots$. Applying this approximated inverse operator to both sides of the inequality yields the desired result. The term in $\mathcal{O}(\lambda_{\max}(\mathbf{F}_B^{-1})^3)$ corresponds to Δ^2 (that is a linear operator involving an order two of \mathbf{F}_B^{-1}) applied to \mathbf{F}_B^{-1} itself. \square

As presented in Theorem 2, neglecting the curvature terms and taking the trace of the inequality is usually valid in practice. Notice that neglecting the curvature terms still yields a different result than the standard Cramér-Rao bound because the error is here defined according to any chosen metric. Hence we can obtain performance bounds for different Riemannian distances (rather than the mean squared error).

Corollary 2 (Second Approximation of intrinsic Bayesian Cramér-Rao lower bound). *Neglecting curvature terms in (40), the expected estimation error is bounded as*

$$\mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}}[d_R^2(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})] \geq \text{Tr}((\mathbb{E}_{\boldsymbol{\theta}}[\mathbf{F}_{\boldsymbol{\theta}}] + \mathbf{F}_{\text{prior}})^{-1}), \quad (43)$$

where $\mathbf{F}_{\boldsymbol{\theta}}$ is defined in (17) and $\mathbf{F}_{\text{prior}}$ is defined as

$$\begin{aligned} (\mathbf{F}_{\text{prior}})_{i,j} &= \mathbb{E}_{\boldsymbol{\theta}}[\text{D} \log f_{\text{prior}}(\boldsymbol{\theta})[\boldsymbol{\Omega}_i] \cdot \text{D} \log f_{\text{prior}}(\boldsymbol{\theta})[\boldsymbol{\Omega}_j]] \\ &= \mathbb{E}_{\boldsymbol{\theta}}[-\text{D}^2 \log f_{\text{prior}}(\boldsymbol{\theta})[\boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j]]. \end{aligned} \quad (44)$$

Proof. We start by removing the curvature terms in Theorem 3, which reads $\mathbf{C} \succeq \mathbf{F}_B^{-1}$. Next, we take the trace of this inequality which yields

$$\mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}}[d_R^2(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})] \geq \text{Tr}(\mathbf{F}_B^{-1}). \quad (45)$$

As for the Euclidean Bayesian case (cf. Equation (6)), since the joint log-likelihood can be expressed as $\log f(\mathbf{y}, \boldsymbol{\theta}) = \log f(\mathbf{y}|\boldsymbol{\theta}) + \log f_{\text{prior}}(\boldsymbol{\theta})$, one can decompose the Bayesian Fisher information matrix as $\mathbf{F}_B = \mathbb{E}_{\boldsymbol{\theta}}[\mathbf{F}_{\boldsymbol{\theta}}] + \mathbf{F}_{\text{prior}}$, which concludes the proof. \square

IV. INTRINSIC BAYESIAN CRAMÉR-RAO BOUNDS FOR COVARIANCE MATRIX ESTIMATION

We consider the fundamental problem of covariance matrix estimation when the data is assumed to follow a centered Gaussian distribution. In the corresponding Bayesian setup, using the conjugate prior consists in assuming that the underlying covariance matrix is drawn from an inverse Wishart distribution. Though this prior distribution has been extensively used in practical estimation problems [41]–[52], there is, to the best of our knowledge, no derivation of a Bayesian performance bound for estimation in this setup (even in the classical framework of Section II). Following from Corollary 2, we obtain these bounds for two performance metrics: the Euclidean one (that yields to a standard performance bound on the mean squared error), and the affine invariant metric for Hermitian positive definite matrices (that yields a bound

on the a distance more naturally suited to the space of covariance matrices). Hence, the two proposed Theorems extend the results of [24] on covariance matrix estimation to a Bayesian context.

A. Data model and MAP (Maximum A Posteriori) and MMSE (Minimum Mean Square Error) estimators

We first present the data model, as well as the associated maximum a posteriori and minimum mean square error estimators: the sample set $\{\mathbf{y}_k\}_{k=1}^n \in \mathbb{C}^p$ is assumed to be i.i.d. and drawn from a Complex Gaussian distribution, denoted $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$, with (unknown) covariance matrix. This covariance matrix belongs to the manifold of Hermitian positive definite matrices, i.e., $\mathbf{\Sigma} \in \mathcal{H}_p^{++}$, where

$$\mathcal{H}_p^{++} = [\mathbf{\Sigma} \in \mathcal{H}_p : \forall \mathbf{x} \in \mathbb{C}^p \setminus \{\mathbf{0}\}, \mathbf{x}^H \mathbf{\Sigma} \mathbf{x} > 0], \quad (46)$$

and where \mathcal{H}_p denotes the set of $p \times p$ Hermitian matrices. The log-likelihood for this sample set is expressed as:

$$\log f(\{\mathbf{y}_k\}_{k=1}^n | \mathbf{\Sigma}) \propto -n \log |\mathbf{\Sigma}| - \sum_{k=1}^n \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{y}_k \mathbf{y}_k^H). \quad (47)$$

Moreover we assume that the covariance matrix is drawn from an inverse Wishart distribution, denoted $\mathbf{\Sigma} \sim \mathcal{IW}((\nu - p)\mathbf{\Sigma}_0, \nu)$, with center $\mathbf{\Sigma}_0$ and ν degrees of freedom (both known). Notice that the choice of scaling in this definition is set so that we have a consistent mean $\mathbb{E}_{\mathbf{\Sigma}}[\mathbf{\Sigma}] = \mathbf{\Sigma}_0$ for any value of $\nu \leq p + 1$. The corresponding probability density function with respect to the Lebesgue measure ($\mu(d\mathbf{\Sigma}) = d\mathbf{\Sigma}$) restricted to \mathcal{H}_p^{++} :

$$\log f_{\mathcal{IW}}(\mathbf{\Sigma} | \mathbf{\Sigma}_0, \nu) \propto -(\nu + p) \log |\mathbf{\Sigma}| - \text{Tr}((\nu - p)\mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0). \quad (48)$$

The choice of this conjugate prior of the Gaussian distribution to model prior knowledge about the covariance matrix has been leveraged in numerous works⁴ [41]–[52]. Given those assumptions, the maximum a posteriori (MAP) estimator of $\mathbf{\Sigma}$ is defined as

$$\hat{\mathbf{\Sigma}}_{\text{MAP-}\mathcal{IW}} = \underset{\mathbf{\Sigma}}{\text{argmax}} \quad f(\{\mathbf{y}_k\}_{k=1}^n | \mathbf{\Sigma}) f_{\mathcal{IW}}(\mathbf{\Sigma} | \mathbf{\Sigma}_0, \nu). \quad (49)$$

The log of the posterior distribution is

$$\begin{aligned} \log f_p &\propto -(\nu + p + n) \log |\mathbf{\Sigma}| - (\nu - p) \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0) \\ &\quad - \sum_{k=1}^n \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{y}_k \mathbf{y}_k^H), \end{aligned} \quad (50)$$

⁴A notable mention is [47] that derives Cramér-Rao bounds for the different problem of estimating the deterministic parameter $\mathbf{\Sigma}_0$ from heterogeneous samples $\mathbf{y}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_k)$ with $\mathbf{\Sigma}_k \sim \mathbf{\Sigma} \sim \mathcal{IW}((\nu - p)\mathbf{\Sigma}_0, \nu)$. The considered Bayesian scenario of estimating the random parameter $\mathbf{\Sigma}$ with prior information about its distribution is more classical, and used in most of the other mentioned references. For this setup, performance bounds were, to the best of our knowledge, not investigated.

from which is easy to obtain a closed form expression for the MAP estimator:

$$\hat{\Sigma}_{MAP-\mathcal{IW}} = \frac{1}{\nu + n + p} ((\nu - p)\Sigma_0 + \sum_{k=1}^n \mathbf{y}_k \mathbf{y}_k^H). \quad (51)$$

The MMSE estimator is given by the posterior mean:

$$\mathbb{E}(\Sigma | \{\mathbf{y}_k\}_{k=1}^n) = \int_{\mathcal{H}^{++}} \Sigma f(\Sigma | \{\mathbf{y}_k\}_{k=1}^n) d\Sigma. \quad (52)$$

From [48], we have:

$$\Sigma \sim \mathcal{IW} \left((\nu - p)\Sigma_0 + \sum_{k=1}^n \mathbf{y}_k \mathbf{y}_k^H, \nu + n \right). \quad (53)$$

Therefore the computation of the mean (52) is direct and gives us:

$$\hat{\Sigma}_{MMSE-\mathcal{IW}} = \frac{1}{\nu + n - p} \left((\nu - p)\Sigma_0 + \sum_{k=1}^n \mathbf{y}_k \mathbf{y}_k^H \right). \quad (54)$$

Finally, we emphasize that \mathcal{H}_p^{++} in (46) is a smooth manifold, as it is an open of the linear space \mathcal{H}_p . Its tangent space at each point Σ is identifiable to $T_\Sigma \mathcal{H}_p^{++} = \mathcal{H}_p$. Endowing this tangent space with a Riemannian metric yields a Riemannian manifold and its corresponding geometry (cf. [35] for a more detailed introduction on this topic). In the following, we study the intrinsic Bayesian Cramér-Rao bound for two of these metrics: the Euclidean metric, and the affine invariant metric on \mathcal{H}_p^{++} .

B. Performance bound based on the Euclidean metric

When endowing \mathcal{H}_p^{++} with the Euclidean metric⁵

$$\langle \Omega_i, \Omega_j \rangle_\Sigma^\mathcal{E} = \Re\{\text{Tr}(\Omega_i \Omega_j)\}, \quad \forall \Omega_i, \Omega_j \in T_\Sigma \mathcal{H}_p^{++}, \quad (55)$$

With this metric, the error vector constructed from the Riemannian logarithm is simply the standard error vector, i.e.:

$$\log_\Sigma^\mathcal{E}(\hat{\Sigma}) = \hat{\Sigma} - \Sigma, \quad (56)$$

and the corresponding error measure is the Euclidean distance

$$d_E^2(\Sigma, \hat{\Sigma}) = \|\hat{\Sigma} - \Sigma\|_F^2. \quad (57)$$

For the upcoming derivations, we first need to define a basis of $T_\Sigma \mathcal{H}_p^{++}$ that is orthonormal with respect to the Euclidean metric (55). In practice, we will use the canonical basis $\{\Omega_i^E\}_{i \in \llbracket 1, p^2 \rrbracket}$ that is indexed over $i \in \llbracket 1, p^2 \rrbracket$ as follows:

⁵When dealing with complex-valued matrices, the real part ensures that the metric defines a proper inner product on \mathcal{H}_p^{++} . As in most works (e.g., [24]), we will however keep this implicit, and omit the notation \Re to lighten the exposition (in particular in the derivations of the appendix).

- For indexes $i \in \llbracket 1, p \rrbracket$, the basis elements $\Omega_i^E \in T_{\Sigma} \mathcal{H}_p^{++}$ refer to the matrices denoted Ω_{ii}^E which are p by p symmetric matrices whose i th diagonal element is one, and zeros elsewhere
- For indexes $i \in \llbracket p+1, p(p+1)/2 \rrbracket$, the basis elements $\Omega_i^E \in T_{\Sigma} \mathcal{H}_p^{++}$ refer to the matrices denoted Ω_{mn}^E which are p by p symmetric matrices whose mn th and nm th elements are both $2^{-1/2}$, and zeros elsewhere.

The mapping between the index i and $(m, n) \in [((1, 2), (1, 3) \dots (1, p), (2, 3), \dots, (p-1, p))]$ is left implicit.

- For indexes $i \in \llbracket p(p+1)/2 + 1, p^2 \rrbracket$, the basis elements $\Omega_i^E \in T_{\Sigma} \mathcal{H}_p^{++}$ refer to the matrices denoted Ω_{mn}^{h-E} which are p by p Hermitian matrices whose mn th element is $2^{-1/2}\sqrt{-1}$, and nm th element is $-2^{-1/2}\sqrt{-1}$, and zeros elsewhere. Again, the second mapping between the index i and $(m, n) \in [((1, 2), (1, 3) \dots (1, p), (2, 3), \dots, (p-1, p))]$ is left implicit.

For the Euclidean distance (57), we obtain the following bound:

Theorem 4 (Bayesian Euclidean Cramér-Rao bound for covariance matrix estimation). *Let $\{\mathbf{y}_k\}_{k=1}^n \in \mathbb{C}^p$ be i.i.d. as $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \Sigma)$, with $\Sigma \sim \mathcal{IW}((\nu - p)\Sigma_0, \nu)$. Let $\hat{\Sigma}$ be an estimator of Σ built from $\{\mathbf{y}_k\}_{k=1}^n \in \mathbb{C}^p$, then*

$$\mathbb{E}_{\mathbf{y}, \Sigma} \left[d_E^2(\Sigma, \hat{\Sigma}) \right] \geq \text{Tr}(\mathbf{F}_{\mathcal{IW}}^{\mathcal{E}})^{-1}, \quad (58)$$

where $\forall (i, j) \in \llbracket 1, p^2 \rrbracket^2$

$$\begin{aligned} (\mathbf{F}_{\mathcal{IW}}^{\mathcal{E}})_{i,j} &= \frac{n\nu^2}{(\nu-p)^2} \text{Tr}(\Sigma_0^{-1} \Omega_i^E \Sigma_0^{-1} \Omega_j^E) \\ &+ \frac{n\nu}{(\nu-p)^2} \text{Tr}(\Sigma_0^{-1} \Omega_i^E) \text{Tr}(\Sigma_0^{-1} \Omega_j^E) + \mathbf{F}_{\text{prior}}^{\mathcal{E}}(i, j) \end{aligned} \quad (59)$$

and

$$(\mathbf{F}_{\text{prior}}^{\mathcal{E}})_{i,j} = \alpha \text{Tr}(\Sigma_0^{-1} \Omega_i^E \Sigma_0^{-1} \Omega_j^E) + \beta \text{Tr}(\Sigma_0^{-1} \Omega_i^E) \text{Tr}(\Sigma_0^{-1} \Omega_j^E), \quad (60)$$

where $\alpha = \frac{\nu^3 + p\nu^2 + 2\nu}{(\nu-p)^2}$ and $\beta = \frac{3\nu^2 - p\nu}{(\nu-p)^2}$.

Proof. cf Section IV-D and appendix A. □

C. Performance bound based on the affine invariant metric

The affine invariant metric on \mathcal{H}_p^{++} is defined as [61], [62] (see footnote 3 concerning the real part operator \Re):

$$\langle \Omega_i, \Omega_j \rangle_{\Sigma}^{\text{AI}} = \Re\{\text{Tr}(\Sigma^{-1} \Omega_i \Sigma^{-1} \Omega_j)\}, \quad \forall \Omega_i, \Omega_j \in T_{\Sigma} \mathcal{H}_p^{++}. \quad (61)$$

Among many other interesting invariance properties, this metric is quadratically dependent on the point Σ^{-1} . This makes the norm of tangent vectors (cf. Figure 1) tend to infinity when the point Σ goes to the

boundaries of \mathcal{H}_p^{++} (when any number of its eigenvalues tend to 0). This Riemannian metric therefore allows us to perceive the boundary of the manifold as being infinitely far, which is more in accordance with the actual geometry of this space than when using the Euclidean metric. The affine invariant metric also appears as the Fisher information metric of the Gaussian model, and the study of the corresponding geometry may reveal unexpected properties of related estimation problems [24]. When endowing \mathcal{H}_p^{++} with the metric (61), the error vector constructed from the Riemannian logarithm is defined as

$$\log_{\Sigma}^{\text{AI}}(\hat{\Sigma}) = \Sigma \log(\Sigma^{-1} \hat{\Sigma}), \quad (62)$$

where \log denotes the standard matrix logarithm. The corresponding error measure is the so-called natural Riemannian distance on \mathcal{H}_p^{++} :

$$d_{\text{AI}}^2(\Sigma, \hat{\Sigma}) = \|\log(\Sigma^{-1} \hat{\Sigma})\|_2^2. \quad (63)$$

For the natural Riemannian distance (63), we obtain the following bound:

Theorem 5 (Intrinsic Bayesian Cramér-Rao bound for covariance matrix estimation). *Let $\{\mathbf{y}_k\}_{k=1}^n \in \mathbb{C}^p$ be i.i.d. as $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \Sigma)$, with $\Sigma \sim \mathcal{IW}((\nu - p)\Sigma_0, \nu)$. Let $\hat{\Sigma}$ be an estimator of Σ built from $\{\mathbf{y}_k\}_{k=1}^n \in \mathbb{C}^p$, then the application of Corollary 2 yields*

$$\mathbb{E}_{\mathbf{y}, \Sigma} [d_{\text{AI}}^2(\Sigma, \hat{\Sigma})] \geq \text{Tr}(\mathbf{F}_{\mathcal{IW}}^{\text{AI}})^{-1}, \quad (64)$$

where

$$\mathbf{F}_{\mathcal{IW}}^{\text{AI}} = n\mathbf{I}_{p^2} + \mathbf{F}_{\text{prior}}^{\text{AI}} \quad (65)$$

and $\forall (i, j) \in \llbracket 1, p^2 \rrbracket^2$

$$\begin{aligned} (\mathbf{F}_{\text{prior}}^{\text{AI}})_{i,j} &= (\nu + p)^2 \delta_{i,j} - (\nu + p) \left((\mathbf{f}_{\text{prior}}^{\text{AI}})_i + (\mathbf{f}_{\text{prior}}^{\text{AI}})_j \right) + (\bar{\mathbf{F}}_{\text{prior}}^{\text{AI}})_{i,j} & \text{if } i, j \leq p \\ (\mathbf{F}_{\text{prior}}^{\text{AI}})_{i,j} &= (\bar{\mathbf{F}}_{\text{prior}}^{\text{AI}})_{i,j} & \text{else} \end{aligned}, \quad (66)$$

with

$$(\mathbf{f}_{\text{prior}}^{\text{AI}})_i = \begin{cases} \nu + p - 2i + 1 & \text{if } i \leq p \\ 0 & \text{if } i > p \end{cases} \quad (67)$$

and

$$(\bar{\mathbf{F}}_{\text{prior}}^{\text{AI}})_{i,j} = \begin{cases} \nu - i + 1 + (\nu + p - 2i + 1)^2 & \text{if } (i, j) \in \llbracket 1, p \rrbracket^2 \text{ and } i = j \\ (\nu + p - 2i + 1)(\nu + p - 2j + 1) & \text{if } (i, j) \in \llbracket 1, p \rrbracket^2 \text{ and } i \neq j \\ \nu + p - 2m + 1 & \text{if } i > p \text{ and } i = j \\ 0 & \text{otherwise} \end{cases} \quad (68)$$

and where m stands for the column index of non-zero element in the element of the basis $\{\Omega_i^E\}_{i=1}^{p^2}$ when identifying $\Omega_i^E = \Omega_{mn}^E$ if $i \in \llbracket p + 1, p(p + 1)/2 \rrbracket$, or $\Omega_i^E = \Omega_{mn}^{h-E}$ if $i \in \llbracket p(p + 1)/2 + 1, p^2 \rrbracket$.

Proof. cf section IV-D and Appendix B. \square

Remark 3. *The main justification for omitting the non-zero curvature terms in the case of the affine invariant metric (i.e. applying Corollary 2 rather than Corollary 1) is that these have been shown to be negligible in the case of deterministic covariance matrix estimation in [24] (which was also observed for other distributions [31]). Ultimately, our simulations also validate this choice because the performance of the estimators is well described by the bound without needing these additional terms.*

To conclude on this result, we notice that, contrarily to the Euclidean case of Theorem 4, the intrinsic bound on the natural Riemannian distance in Theorem 5 does not depend on the hyperparameter Σ_0 , but only on the dimensions n , p , and the degrees of freedom ν . This interesting property offers an easier interpretation of the bound, but will also have a more fundamental impact (the possibility of statistical efficiency regarding the chosen metric), as illustrated in the next section.

D. Proofs outline

Independently of the chosen metric, we consider that we have an orthonormal basis of the tangent space that is denoted $\{\Omega_i\}_{i=1}^{p^2}$ (the canonical one for the Euclidean metric and a more complex for the affine invariant metric which is developed in appendix B). We then have to compute the matrix⁶

$$\mathbf{F}_{\mathcal{IW}} = \mathbb{E} [\mathbf{F}_{\Sigma}(\Sigma)] + \mathbf{F}_{\text{prior}}, \quad (69)$$

where

$$(\mathbf{F}_{\Sigma}(\Sigma))_{i,j} = n \text{Tr} (\Sigma^{-1} \Omega_i \Sigma^{-1} \Omega_j) \quad (70)$$

is the Fisher information matrix of the Gaussian model (cf. [24], [31], [35] for its derivation with matching notations). The second term $\mathbf{F}_{\text{prior}}$ is defined as

$$(\mathbf{F}_{\text{prior}})_{i,j} = \mathbb{E} [D_{\Sigma} \log f_{\mathcal{IW}}(\Sigma)[\Omega_i] \cdot D_{\Sigma} \log f_{\mathcal{IW}}(\Sigma)[\Omega_j]]. \quad (71)$$

Given the directional derivatives

$$\begin{aligned} D_{\Sigma} (\log |\Sigma|) [\xi] &= \text{Tr}(\Sigma^{-1} \xi) \\ D_{\Sigma} (\text{Tr}(\Sigma^{-1} \Sigma_0)) [\xi] &= -\text{Tr}(\Sigma^{-1} \xi \Sigma^{-1} \Sigma_0). \end{aligned} \quad (72)$$

We have:

$$\begin{aligned} D_{\Sigma} \log f_{\mathcal{IW}}(\Sigma)[\Omega_i] &= -(\nu + p) \text{Tr}(\Sigma^{-1} \Omega_i) \\ &+ (\nu - p) \text{Tr}(\Sigma^{-1} \Omega_i \Sigma^{-1} \Sigma_0). \end{aligned} \quad (73)$$

⁶From now, all expectations are only taken over Σ , but denoted \mathbb{E} to lighten the exposition.

Hence, we have

$$\begin{aligned}
(\mathbf{F}_{\text{prior}})_{i,j} &= (\nu + p)^2 \mathbb{E} [\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i) \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_j)] \\
&+ (\nu - p)^2 \mathbb{E} [\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i \mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0) \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_j \mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0)] \\
&- (\nu^2 - p^2) \mathbb{E} [\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i) \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_j \mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0)] \\
&- (\nu^2 - p^2) \mathbb{E} [\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_j) \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i \mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0)].
\end{aligned} \tag{74}$$

Hence, the computation of $\mathbf{F}_{\text{prior}}$ and $\mathbb{E}[\mathbf{F}_{\Sigma}(\mathbf{\Sigma})]$ involves the calculus of four core expectations:

$$\begin{aligned}
T_1(i, j) &= \mathbb{E} [\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i) \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_j)] \\
T_2(i, j) &= \mathbb{E} [\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i \mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0) \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_j \mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0)] \\
T_3(i, j) &= \mathbb{E} [\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_j) \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i \mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0)] \\
T_4(i, j) &= \mathbb{E} [\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i \mathbf{\Sigma}^{-1} \mathbf{\Omega}_j)]
\end{aligned} \tag{75}$$

At this point, the proof of the two Theorems differ because we need to define a basis of the tangent space $\{\mathbf{\Omega}_i\}_{i=1}^{p^2}$ that is orthogonal with respect to the chosen error metric:

- For the Euclidean metric, we can use the standard canonical basis of \mathcal{H}_p presented in section IV-B. The computation of the four expectations then requires to derive new results about traces of functions of inverse Wishart matrices. Full derivations are provided in Appendix A.
- The affine invariant metric requires a more careful construction of the basis thanks to the Bartlett decomposition, in order for the expectations to be actually tractable. The detailed derivation are provided in Appendix B.

These tedious derivations being achieved, we have all the necessary objects to obtain the final inequalities: the results are obtained by applying Corollary 2 and the relation (16). As in [24] the curvature terms are neglected in the case of the affine invariant metric.

V. NUMERICAL EXPERIMENTS

A. Parameters and method

We consider a Toeplitz matrix $(\mathbf{\Sigma}_0)_{i,j} = \rho^{|i-j|}$ where $\rho = 0.5$. The data size is $p = 5$. The degrees of freedom for the \mathcal{IW} prior is denoted ν and will take two values $\nu = 40, 100$. The number of trials for estimating the MSE (Mean Square Error) is equal to 1000. We sample the data by using $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_{\mathcal{IW}}$. We compute the SCM, the MAP (51) and the MMSE estimators (54). And finally, we evaluate the MSE either by using the Euclidean distance $d_{\mathcal{E}}^2$ for the Euclidean inequality or the natural distance $d_{\mathcal{AI}}^2$ for the intrinsic inequality. The Euclidean and intrinsic Cramér-Rao bounds are computed by using the formulas from theorems 4 and 5. The code is available at the following address https://github.com/flbouchard/intrinsic_Bayesian_CRB.

B. Results

Figures 4a and 4b show the MSE computed either with the Euclidean distance or the natural distance and the corresponding Cramér-Rao bounds (CRB). We retrieve the classical result in the Euclidean case where the SCM is proven to be efficiency. On the opposite, this efficiency is lost when we consider the natural properties of a covariance matrix. This result is also known and firstly reported in [24] and shows the interest of the intrinsic bound in demonstrating properties of an estimator that the classical Euclidean bound fails to exhibit.

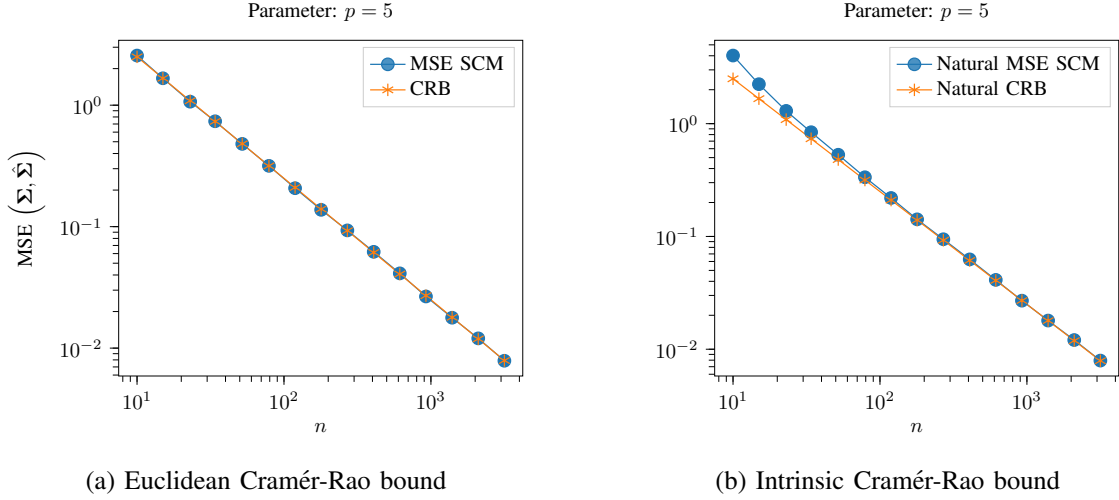


Fig. 4: Euclidean Cramér-Rao bound and MSE (left) and intrinsic Cramér-Rao bound and distance w.r.t. n . The data size is $p = 5$.

Now, we turn on the study of the MAP and MMSE estimators as well as the corresponding Bayesian Cramér-Rao bounds (BCRB). In particular, we also show the asymptotic Bayesian bounds which are computed without considering the term \mathbf{F}_{prior} (i.e., only $\text{Tr}(\mathbb{E}_{\Sigma} [\mathbf{F}_{\Sigma}]^{-1})$) in (59) and (65). We consider two cases: $\nu = 40$ and $\nu = 100$. The results are shown in Figures 5a and 5b for the first value of ν and in Figures 6a and 6b for the second one. For both values, conclusions are similar. We have the property of non efficiency for both the MAP and MMSE estimators in the Euclidean study. Concerning the MAP estimator, this result was expected from [9] p.7 since the expression of the classical (non-Bayesian) Fisher information matrix $\mathbf{F}_{\Sigma}(\Sigma)$ depends on Σ . In the Riemannian framework, the asymptotic efficiency is achieved, which was also expected for the MAP estimator for the same aforementioned reason since, in this case, the intrinsic non-Bayesian Fisher information matrix is proven to be independent of the studied parameter Σ . This intrinsic analysis allows to conclude that the MAP and MMSE estimators are valid

estimators as soon as we have enough samples to estimate them (depending on ν). As expected, the MMSE estimator performs slightly better than the MAP in particular for a low sample support.

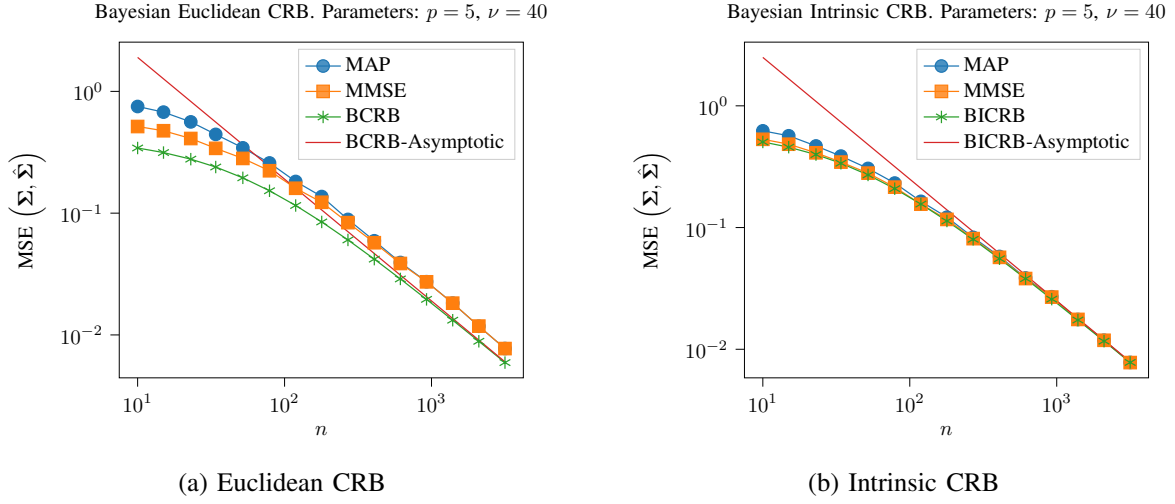


Fig. 5: Euclidean Bayesian Cramér-Rao bound and MSE (left) and intrinsic Bayesian Cramér-Rao bound and expectation of the natural distance w.r.t. n . The data size is $p = 5$ and the number of degrees of freedom of the \mathcal{IW} prior is $\nu = 40$.

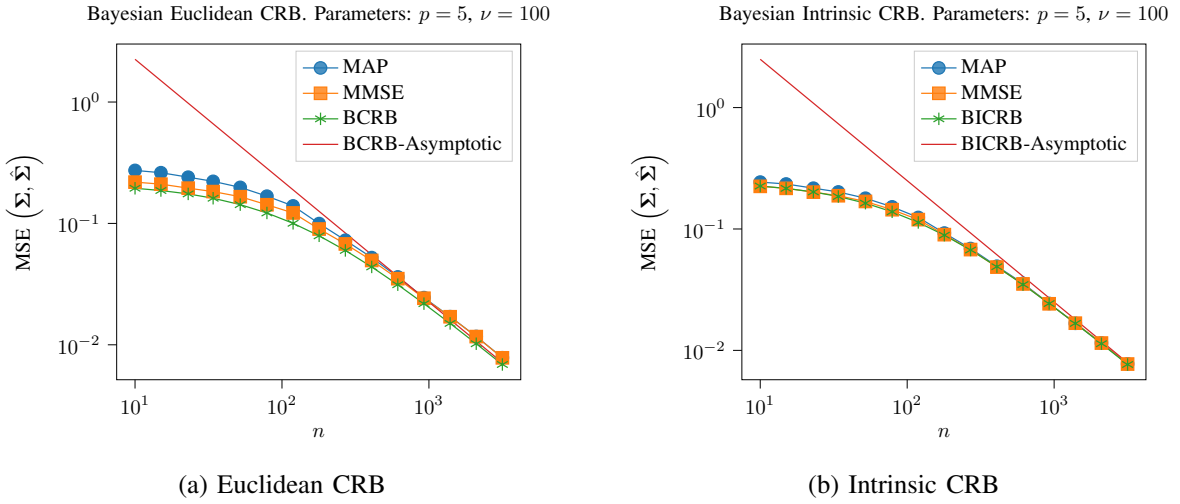


Fig. 6: Euclidean Bayesian Cramér-Rao bound and MSE (left) and intrinsic Bayesian Cramér-Rao bound and expectation of the natural distance w.r.t. n . The data size is $p = 5$ and the number of degrees of freedom of the \mathcal{IW} prior is $\nu = 100$.

VI. CONCLUSION

In the context of Bayesian estimation, when the parameter to estimate lies in a manifold, we have proposed a new intrinsic Van Trees inequality between a covariance of the estimation error defined with geometric tools and an intrinsic Bayesian Fisher information. This derivation is made by using some assumptions on the manifold of interest and the prior distribution. We illustrated this result by considering the problem of covariance estimation when the data follow a Gaussian distribution and the prior distribution is an inverse Wishart. Numerical simulation led to interesting conclusions on the MAP and the MMSE estimators which seem to be asymptotic efficiency in the natural inequality which is not the case when the study is made with the Euclidean formalism.

REFERENCES

- [1] H. Cramér, *Mathematical Methods of Statistics*, New York: Princeton University, 1946.
- [2] C. Radhakrishna Rao, *Information and the Accuracy Attainable in the Estimation of Statistical Parameters*, pp. 235–247, Springer New York, New York, NY, 1992.
- [3] A. Bhattacharyya, “On some analogues of the amount of information and their use in statistical estimation,” *Sankhya: The Indian Journal of Statistics (1933-1960)*, vol. 8, no. 1, pp. 1–14, 1946.
- [4] E. W. Barankin, “Locally Best Unbiased Estimates,” *The Annals of Mathematical Statistics*, vol. 20, no. 4, pp. 477 – 501, 1949.
- [5] Koby Todros and Joseph Tabrikian, “General classes of performance lower bounds for parameter estimation—part ii: Bayesian bounds,” *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5064–5082, 2010.
- [6] Eric Chaumette, Jérôme Galy, Angela Quinlan, and Pascal Larzabal, “A new Barankin bound approximation for the prediction of the threshold region performance of maximum likelihood estimators,” *IEEE Transactions on Signal Processing*, vol. 56, no. 11, pp. 5319–5333, 2008.
- [7] Tirza Routtenberg and Joseph Tabrikian, “Cyclic Barankin-type bounds for non-Bayesian periodic parameter estimation,” *IEEE Transactions on Signal Processing*, vol. 62, no. 13, pp. 3321–3336, 2014.
- [8] H.L. Van Trees, *Estimation and modulation theory*, vol. 1, John Wiley and Sons, 2001.
- [9] H.L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, John Wiley and Sons, 2007.
- [10] E. Weinstein and A. J. Weiss, “A general class of lower bounds in parameter estimation,” *IEEE Trans. Inform. Theory*, vol. 34, pp. 338–342, 1988.
- [11] Eric Chaumette, Alexandre Renaux, and Mohammed Nabil El Korso, “A class of Weiss–Weinstein bounds and its relationship with the Bobrovsky–Mayer–Wolf–Zakai bounds,” *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2226–2240, 2017.
- [12] J.D. Gorman and A.O. Hero, “Lower bounds for parametric estimation with constraints,” *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1285–1301, 1990.
- [13] Petre Stoica and Boon Chong Ng, “On the Cramér–Rao bound under parametric constraints,” *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 177–179, 1998.
- [14] Thomas L Marzetta, “A simple derivation of the constrained multiple parameter Cramer-Rao bound,” *IEEE Transactions on Signal Processing*, vol. 41, no. 6, pp. 2247–2249, 1993.

- [15] Zvika Ben-Haim and Yonina C Eldar, “On the constrained Cramér–Rao bound with a singular Fisher information matrix,” *IEEE Signal Processing Letters*, vol. 16, no. 6, pp. 453–456, 2009.
- [16] A Al-Sharadqah and KC Ho, “Constrained cramér–Rao lower bound in errors-in variables (eiv) models: Revisited,” *Statistics & Probability Letters*, vol. 135, pp. 118–126, 2018.
- [17] Eyal Nitzan, Tirza Routtenberg, and Joseph Tabrikian, “Cramér–Rao bound for constrained parameter estimation using lehmann-unbiasedness,” *IEEE Transactions on Signal Processing*, vol. 67, no. 3, pp. 753–768, 2018.
- [18] Eyal Nitzan, Tirza Routtenberg, and Joseph Tabrikian, “Cramér–Rao bound under norm constraint,” *IEEE Signal Processing Letters*, vol. 26, no. 9, pp. 1393–1397, 2019.
- [19] Chengfang Ren, Julien Le Kernec, Jerome Galy, Eric Chaumette, Pascal Larzabal, and Alexandre Renaux, “A constrained hybrid Cramér–Rao bound for parameter estimation,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 3472–3476.
- [20] Eyal Nitzan, Tirza Routtenberg, and Joseph Tabrikian, “Barankin-type bound for constrained parameter estimation,” *IEEE Transactions on Signal Processing*, 2023.
- [21] Leif Erik Andersson, Lars Imsland, Edmund Fjørland Brekke, and Francesco Scibilia, “Constrained posterior Cramér–Rao bound for discrete-time systems,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 3768–3773, 2017.
- [22] Clémence Prévost, Eric Chaumette, Konstantin Usevich, David Brie, and Pierre Comon, “On Cramér–Rao lower bounds with random equality constraints,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5355–5359.
- [23] Harrie Hendriks, “A Cramér–Rao type lower bound for estimators with values in a manifold,” *Journal of Multivariate Analysis*, vol. 38, no. 2, pp. 245–261, 1991.
- [24] Steven T. Smith, “Covariance, subspace, and intrinsic Cramér–Rao bounds,” *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, 2005.
- [25] J. Xavier and V. Barroso, “Intrinsic variance lower bound (ivlb): an extension of the Cramér–Rao bound to Riemannian manifolds,” in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, 2005, vol. 5, pp. v/1033–v/1036 Vol. 5.
- [26] Axel Barrau and Silvere Bonnabel, “A note on the intrinsic Cramér–Rao bound,” in *International Conference on Geometric Science of Information*. Springer, 2013, pp. 377–386.
- [27] Nicolas Boumal, “On intrinsic Cramér–Rao bounds for Riemannian submanifolds and quotient manifolds,” *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1809–1821, 2013.
- [28] Nicolas Boumal, Amit Singer, P-A Absil, and Vincent D Blondel, “Cramér–Rao bounds for synchronization of rotations,” *Information and Inference: A Journal of the IMA*, vol. 3, no. 1, pp. 1–39, 2014.
- [29] Nicolas Boumal, *Optimization and estimation on manifolds*, Ph.D. thesis, Université catholique de Louvain, feb 2014.
- [30] S. Bonnabel and A. Barrau, “An intrinsic Cramér–Rao bound on Lie groups,” in *International Conference on Networked Geometric Science of Information*, 2015, pp. 664–672.
- [31] Arnaud Breloy, Guillaume Ginolhac, Alexandre Renaux, and Florent Bouchard, “Intrinsic Cramér–Rao bounds for scatter and shape matrices estimation in CES distributions,” *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 262–266, 2019.
- [32] Victor Solo and Gregory S. Chirikjian, “On the Cramér–Rao bound in riemannian manifolds with application to $SO(3)$,” in *2020 59th IEEE Conference on Decision and Control (CDC)*, 2020, pp. 4117–4122.
- [33] S. Labsir, D. Medina, J. Vilà-Valls, and E. Chaumette, “An intrinsic McAulay–Seidman bound for parameters evolving on matrix Lie groups,” in *Proceedings of the 56th Asilomar Conference on Signals, Systems and Computers*, 2022.

- [34] H. W. M. Hendriks, J. H. M. Janssen, and F. H. Ruymgaart, “A Cramér-Rao type inequality for random variables in Euclidean manifolds,” *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)*, vol. 54, no. 3, pp. 387–401, 1992.
- [35] Florent Bouchard, Arnaud Breloy, Antoine Collas, Alexandre Renaux, and Guillaume Ginolhac, “The Fisher-Rao geometry of CES distributions,” *arXiv preprint arXiv:2310.01032*, 2023.
- [36] P.E. Jupp, “A Van Trees inequality for estimators on manifolds,” *Journal of Multivariate Analysis*, vol. 101, no. 8, pp. 1814–1825, 2010.
- [37] M. Ashok Kumar and Kumar Vijay Mishra, “Information geometric approach to Bayesian lower error bounds,” 2018.
- [38] Shun-ichi Amari and Hiroshi Nagaoka, *Methods of information geometry*, vol. 191, American Mathematical Soc., 2000.
- [39] Samy Labsir, Audrey Giremus, B Yver, and Thomas Benoudiba-Campanini, “An intrinsic Bayesian bound for estimators on the Lie groups SO (3) and SE (3),” *Signal Processing*, vol. 214, pp. 109232, 2024.
- [40] Samy Labsir, Alexandre Renaux, Jordi Vilà-Valls, and Eric Chaumette, “Barankin, Mcaulay–Seidman and Cramér–Rao bounds on matrix Lie groups,” *Automatica*, vol. 156, pp. 111199, 2023.
- [41] John Barnard, Robert McCulloch, and Xiao-Li Meng, “Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage,” *Statistica Sinica*, pp. 1281–1311, 2000.
- [42] M. Bouriga and O. Féron, “Estimation of covariance matrices based on hierarchical inverse-Wishart priors,” *Journal of Statistical Planning and Inference*, vol. 143, no. 4, pp. 795–808, 2013.
- [43] C. Fraley and A. Raftery, “Bayesian regularization for normal mixture estimation and model-based clustering,” *Journal of Classification*, vol. 24, pp. 155–181, 2007.
- [44] Mohsen Pourahmadi, “Covariance Estimation: The GLM and Regularization Perspectives,” *Statistical Science*, vol. 26, no. 3, pp. 369 – 387, 2011.
- [45] Joseph R Guerci and Edward J Baranoski, “Knowledge-aided adaptive radar at DARPA: An overview,” *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 41–50, 2006.
- [46] Olivier Besson, Jean-Yves Tournet, and Stphanie Bidon, “Knowledge-aided Bayesian detection in heterogeneous environments,” *IEEE Signal Processing Letters*, vol. 14, no. 5, pp. 355–358, 2007.
- [47] Olivier Besson, Stéphanie Bidon, and Jean-Yves Tournet, “Bounds for estimation of covariance matrices from heterogeneous samples,” *IEEE transactions on signal processing*, vol. 56, no. 7, pp. 3357–3362, 2008.
- [48] Francesco Bandiera, Olivier Besson, and Giuseppe Ricci, “Knowledge-aided Bayesian covariance matrix estimation in compound-Gaussian clutter,” in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 2574–2577.
- [49] Francesco Bandiera, Olivier Besson, and Giuseppe Ricci, “Knowledge-aided covariance matrix estimation and adaptive detection in compound-Gaussian noise,” *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5391–5396, 2010.
- [50] Antonio De Maio, A Farina, and G Foglia, “Knowledge-aided Bayesian radar detectors & their application to live data,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 170–183, 2010.
- [51] Stephanie Bidon, Olivier Besson, and Jean-Yves Tournet, “Knowledge-aided STAP in heterogeneous clutter using a hierarchical Bayesian algorithm,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 3, pp. 1863–1879, 2011.
- [52] Yikai Wang, Wei Xia, and Zishu He, “CFAR knowledge-aided radar detection with heterogeneous samples,” *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 693–697, 2017.
- [53] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian geometry*, Springer, 1990.
- [54] S. Lang, *Differential and Riemannian manifolds*, Springer, 2012.
- [55] J. M. Lee, *Riemannian manifolds: an introduction to curvature*, Springer, 2006.

- [56] Pierre-Antoine Absil, Robert Mahony, and Rodolphe Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.
- [57] Nicolas Boumal, *An introduction to optimization on smooth manifolds*, Cambridge University Press, 2023.
- [58] R Ben Abdallah, Arnaud Breloy, Mohammed Nabil El Korso, and David Lautru, “Bayesian signal subspace estimation with compound gaussian sources,” *Signal Processing*, vol. 167, pp. 107310, 2020.
- [59] Salem Said, Lionel Bombrun, Yannick Berthoumieu, and Jonathan H Manton, “Riemannian Gaussian distributions on the space of symmetric positive definite matrices,” *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2153–2170, 2017.
- [60] Nicolas Boumal, “On intrinsic cramer-rao bounds for riemannian submanifolds and quotient manifolds,” *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1809–1821, 2013.
- [61] Rajendra Bhatia, *Positive definite matrices*, Princeton university press, 2009.
- [62] L. T. Skovgaard, “A Riemannian geometry of the multivariate normal model,” *Scandinavian journal of statistics*, pp. 211–223, 1984.
- [63] D. Maiwald and D. Kraus, “Calculations of moments of complex Wishart and inverse Whishart distributed matrices,” *IEE Proc. Radar Sonar Navig.*, vol. 147, no. 4, pp. 162–168, August 2000.

APPENDIX A

EXPECTATIONS 75 FOR THEOREM 4 (EUCLIDEAN METRIC)

This proof is based on the computation of expectations of traces of functions of Wishart matrices. Main results have been derived in [63], however they do not cover the higher orders that we have to deal with here. The required results are obtained thanks to the following lemma:

Lemma 1. *Let $S \sim \mathcal{W}(K, \frac{1}{K}\Sigma)$ which follows a Wishart distribution of scale matrix Σ and degrees of freedom K , then we have the following expectations*

$$\begin{aligned}
 \mathbb{E}[\text{Tr}(\mathbf{A}\mathbf{S}\mathbf{B}\mathbf{S}) \text{Tr}(\mathbf{C}\mathbf{S})] &= \text{Tr}(\mathbf{A}\Sigma\mathbf{B}\Sigma) \text{Tr}(\mathbf{C}\Sigma) \\
 &+ \frac{1}{K}(\text{Tr}(\mathbf{A}\Sigma\mathbf{B}\Sigma\mathbf{C}\Sigma) + \\
 &\text{Tr}(\mathbf{A}\Sigma\mathbf{C}\Sigma\mathbf{B}\Sigma) + \text{Tr}(\mathbf{A}\Sigma) \text{Tr}(\mathbf{B}\Sigma) \text{Tr}(\mathbf{C}\Sigma)) \\
 &+ \frac{1}{K^2}(\text{Tr}(\mathbf{A}\Sigma\mathbf{C}\Sigma) \text{Tr}(\mathbf{B}\Sigma) + \text{Tr}(\mathbf{B}\Sigma\mathbf{C}\Sigma) \text{Tr}(\mathbf{A}\Sigma)),
 \end{aligned} \tag{76}$$

and

$$\begin{aligned}
& \mathbb{E}[\text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}) \text{Tr}(\mathbf{C}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma})] = \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}) \text{Tr}(\mathbf{C}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}) \\
& + \frac{1}{K} (\text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}) + \\
& \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}) + \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}) \\
& + \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}) + \\
& \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}) \text{Tr}(\mathbf{C}\mathbf{\Sigma}) \text{Tr}(\mathbf{D}\mathbf{\Sigma}) + \\
& \text{Tr}(\mathbf{C}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}) \text{Tr}(\mathbf{A}\mathbf{\Sigma}) \text{Tr}(\mathbf{B}\mathbf{\Sigma})) \\
& + \frac{1}{K^2} (\text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}) \text{Tr}(\mathbf{C}\mathbf{\Sigma}) + \\
& \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}) \text{Tr}(\mathbf{D}\mathbf{\Sigma}) + \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}) \text{Tr}(\mathbf{B}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}) \\
& + \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}) \text{Tr}(\mathbf{C}\mathbf{\Sigma}) + \\
& \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}) \text{Tr}(\mathbf{D}\mathbf{\Sigma}) + \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}) \text{Tr}(\mathbf{B}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}) \\
& + \text{Tr}(\mathbf{A}\mathbf{\Sigma}) \text{Tr}(\mathbf{B}\mathbf{\Sigma}) \text{Tr}(\mathbf{C}\mathbf{\Sigma}) \text{Tr}(\mathbf{D}\mathbf{\Sigma}) + \\
& \text{Tr}(\mathbf{B}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}) \text{Tr}(\mathbf{A}\mathbf{\Sigma}) + \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}) \text{Tr}(\mathbf{B}\mathbf{\Sigma}) \\
& + \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}) \text{Tr}(\mathbf{B}\mathbf{\Sigma}) + \text{Tr}(\mathbf{B}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}) \text{Tr}(\mathbf{A}\mathbf{\Sigma})) \\
& + \frac{1}{K^3} (\text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}) + \\
& \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}) \text{Tr}(\mathbf{B}\mathbf{\Sigma}) \text{Tr}(\mathbf{D}\mathbf{\Sigma}) + \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}) \text{Tr}(\mathbf{B}\mathbf{\Sigma}) \text{Tr}(\mathbf{C}\mathbf{\Sigma}) \\
& + \text{Tr}(\mathbf{B}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}) \text{Tr}(\mathbf{A}\mathbf{\Sigma}) \text{Tr}(\mathbf{D}\mathbf{\Sigma}) + \text{Tr}(\mathbf{B}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}) \text{Tr}(\mathbf{A}\mathbf{\Sigma}) \text{Tr}(\mathbf{C}\mathbf{\Sigma}) + \\
& \text{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}\mathbf{D}\mathbf{\Sigma}))
\end{aligned} \tag{77}$$

Proof. The derivations are very long and are therefore omitted in this paper. Nevertheless, they can be retrieved by using the method of [63] (in particular the relation $\text{Tr}(\mathbf{A}_{(1)} \dots \mathbf{A}_{(n)}) = \sum_{i_1, \dots, i_n} a_{i_1 i_2}^{(1)} \dots a_{i_n i_1}^{(n)}$) but considering a higher order than in this reference. \square

Let go back to the proof of Theorem 4. We recall that we have to compute the following expectations:

- $T_1(i, j) = \mathbb{E} \left[\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i^E) \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_j^E) \right],$
- $T_2(i, j) = \mathbb{E} \left[\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i^E \mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0) \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_j^E \mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0) \right],$
- $T_3(i, j) = \mathbb{E} \left[\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_j^E) \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i^E \mathbf{\Sigma}^{-1} \mathbf{\Sigma}_0) \right],$
- $T_4(i, j) = \mathbb{E} \left[\text{Tr} \left[\mathbf{\Sigma}^{-1} \mathbf{\Omega}_i^E \mathbf{\Sigma}^{-1} \mathbf{\Omega}_j^E \right] \right].$

Since $\mathbf{\Sigma} \sim \mathcal{IW}(\nu, (\nu - p)\mathbf{\Sigma}_0)$, it is easy to notice that $\mathbf{\Sigma}^{-1} \sim \mathcal{W}(\nu, \frac{\mathbf{\Sigma}_0^{-1}}{\nu - p})$. Therefore, we can apply the results of Lemma 1, but taking into account the scale of the problem considered in our case.

Let us introduce the changes of variables $\tilde{\mathbf{\Sigma}}^{-1} = \frac{1}{\nu} \mathbf{\Sigma}^{-1}$ and $\tilde{\mathbf{\Sigma}}_0^{-1} = \frac{1}{\nu - p} \mathbf{\Sigma}_0^{-1}$. In this case, we have $\tilde{\mathbf{\Sigma}}^{-1} \sim \mathcal{W}(\nu, \frac{\tilde{\mathbf{\Sigma}}_0^{-1}}{\nu})$. We are then in the conditions of the Lemma 1 which can be applied to each term

involving Σ^{-1} in (75), then properly re-scaled by $\nu/(\nu - p)$ to have the final expectations. We finally obtain for the 4 terms:

$$\begin{aligned}
T_1(i, j) &= \frac{1}{(\nu-p)^2} (\nu^2 A + \nu B) \\
T_2(i, j) &= \frac{1}{(\nu-p)^4} (\nu^4 A + \nu^3 (4B + 2pA) \\
&\quad + \nu^2 (5A + 5pB + p^2 A) + \nu (3pA + 2B + p^2 B)) , \\
T_3(i, j) &= \frac{1}{(\nu-p)^3} (\nu^3 A + \nu^2 (2B + pA) \nu (pB + A)) \\
T_4(i, j) &= \frac{1}{(\nu-p)^2} (\nu^2 B + \nu A)
\end{aligned} \tag{78}$$

where $A = \text{Tr}(\Sigma_0^{-1} \Omega_i^E) \text{Tr}(\Sigma_0^{-1} \Omega_j^E)$ and $B = \text{Tr}(\Sigma_0^{-1} \Omega_i^E \Sigma_0^{-1} \Omega_j^E)$. Finally, after basic manipulations, we have the following result:

$$\begin{aligned}
\left(\mathbf{F}_{\text{prior}}^{\mathcal{E}} \right)_{i,j} &= \alpha \text{Tr}(\Sigma_0^{-1} \Omega_i^E \Sigma_0^{-1} \Omega_j^E) \\
&\quad + \beta \text{Tr}(\Sigma_0^{-1} \Omega_i^E) \text{Tr}(\Sigma_0^{-1} \Omega_j^E),
\end{aligned} \tag{79}$$

where $\alpha = \frac{\nu^3 + p\nu^2 + 2\nu}{(\nu-p)^2}$ and $\beta = \frac{3\nu^2 - p\nu}{(\nu-p)^2}$. By using again the term $T_4(i, j)$, the final result is factorized as

$$\begin{aligned}
\left(\mathbf{F}_{\mathcal{IW}}^{\mathcal{E}} \right)_{i,j} &= \frac{n\nu^2}{(\nu-p)^2} \text{Tr}(\Sigma_0^{-1} \Omega_i^E \Sigma_0^{-1} \Omega_j^E) \\
&\quad + \frac{n\nu}{(\nu-p)^2} \text{Tr}(\Sigma_0^{-1} \Omega_i^E) \text{Tr}(\Sigma_0^{-1} \Omega_j^E) + \left(\mathbf{F}_{\text{prior}}^{\mathcal{E}} \right)_{i,j}.
\end{aligned} \tag{80}$$

APPENDIX B

EXPECTATIONS 75 FOR THEOREM 5 (AFFINE INVARIANT METRIC)

This proof is decomposed in three steps. Firstly, we have to choose a basis which has been to be orthonormal w.r.t. the affine invariant metric (61). Secondly, from this chosen basis, we reduce the terms T_1 , T_2 and T_3 of (75) (we will show that the computation of T_4 is not needed for this bound). Finally, the expectations for these reduced terms are computed.

A. Choice of the basis

In this section, we have to make a choice for the basis $\{\Omega_i\}$. This basis has been to be orthonormal w.r.t. the natural inner product $n \text{Tr} [\Sigma^{-1} \Omega_i \Sigma^{-1} \Omega_j]$. As for the Euclidean proof, since $\Sigma \sim \mathcal{IW}(\nu, (\nu-p)\Sigma_0)$, we have:

$$\Sigma^{-1} \sim \mathcal{W}(\nu, \frac{1}{\nu-p} \Sigma_0^{-1}). \tag{81}$$

Let us consider the Cholesky decomposition of Σ_0^{-1} :

$$\Sigma_0^{-1} = \mathbf{L} \mathbf{L}^H, \tag{82}$$

where \mathbf{L} is a lower triangular matrix. We also use the Bartlett decomposition of Σ^{-1} :

$$\Sigma^{-1} = \frac{1}{\nu-p} \mathbf{L} \mathbf{A} \mathbf{A}^H \mathbf{L}^H, \tag{83}$$

where \mathbf{A} is also a lower triangular matrix and where all the elements are independent random variables:

$$\begin{aligned} a_{i,j} &\sim \mathcal{CN}(0, 1), i > j \\ a_{i,i}^2 &\sim \frac{1}{2}\chi_{2(\nu-i+1)}^2 \end{aligned} \quad (84)$$

Let us define a square root of Σ :

$$\Sigma = \mathbf{H}\mathbf{H}^H. \quad (85)$$

We know there exists an unitary matrix $\mathbf{U} \in \mathcal{U}_p$ such that⁷:

$$\mathbf{H} = \Sigma^{1/2}\mathbf{U}. \quad (86)$$

Based on this definition of the square root, we choose to build the natural basis from the Euclidean basis and \mathbf{H} :

$$\Omega_i = \mathbf{H}\Omega_i^E\mathbf{H}^H. \quad (87)$$

Let us choose now a matrix \mathbf{H} . From the Bartlett decomposition (83):

$$\Sigma = (\nu - p)\mathbf{L}^{-H}\mathbf{A}^{-H}\mathbf{A}^{-1}\mathbf{L}^{-1}. \quad (88)$$

Therefore, we choose

$$\mathbf{H} = \sqrt{\nu - p}\mathbf{L}^{-H}\mathbf{A}^{-H}. \quad (89)$$

It is easy to see that the basis defined in (87) with \mathbf{H} given as in (89) is orthonormal w.r.t. the natural inner product.

B. Computation of the terms T_1 , T_2 and T_3 of (75)

We recall that we have to compute the matrix

$$\mathbf{F}_{\mathcal{TW}}^{AI} = \mathbb{E}_{\Sigma} [\mathbf{F}_{\Sigma}(\Sigma)] + \mathbf{F}_{\text{prior}}^{AI}, \quad (90)$$

where $\mathbf{F}_{\Sigma}(\Sigma) = n\mathbf{I}_p$ [31] and

$$\begin{aligned} \left(\mathbf{F}_{\text{prior}}^{AI}\right)_{i,j} &= [(\nu + p)^2 T_1(i, j) + (\nu - p)^2 T_2(i, j) \\ &\quad - (\nu^2 - p^2) T_3(i, j) - (\nu^2 - p^2) T_3(j, i)]. \end{aligned} \quad (91)$$

We notice, compared to the Euclidean bound, that we only have to compute T_1 , T_2 and T_3 since $\mathbf{F}_{\Sigma}(\Sigma)$ does not involve Σ .

⁷ \mathcal{U}_p is the set of unitary matrices of size $p \times p$

Let us start to compute the terms $\text{Tr}(\Sigma^{-1}\Omega_i)$ and $\text{Tr}(\Sigma^{-1}\Omega_i\Sigma^{-1}\Sigma_0)$ by using the basis chosen in the previous section. For the first term, we have:

$$\begin{aligned}\text{Tr}(\Sigma^{-1}\Omega_i) &= \text{Tr}(\Sigma^{-1}\mathbf{H}\Omega_i^E\mathbf{H}^H) \\ &= \text{Tr}\left(\frac{1}{\nu-p}\mathbf{L}\mathbf{A}\mathbf{A}^H\mathbf{L}^H\mathbf{L}^{-H}\mathbf{A}^{-H}\sqrt{\nu-p}\Omega_i^E\sqrt{\nu-p}\mathbf{A}^{-1}\mathbf{L}^{-1}\right) \\ &= \text{Tr}(\Omega_i^E).\end{aligned}\tag{92}$$

The second term is

$$\begin{aligned}\text{Tr}(\Sigma^{-1}\Omega_i\Sigma^{-1}\Sigma_0) &= \text{Tr}((\nu-p-1)\Sigma^{-1}\mathbf{L}^{-H}\mathbf{A}^{-H}\Omega_i^E\mathbf{A}^{-1}\mathbf{L}^{-1}\Sigma^{-1}\Sigma_0) \\ &= \text{Tr}\left(\frac{\nu-p}{\nu-p}\Sigma^{-1}\mathbf{L}^{-H}\mathbf{A}^{-H}\Omega_i^E\mathbf{A}^{-1}\mathbf{L}^{-1}\Sigma^{-1}\mathbf{L}^{-H}\mathbf{L}^{-1}\right) \\ &= \text{Tr}\left(\frac{1}{\nu-p}\mathbf{L}\mathbf{A}\mathbf{A}^H\mathbf{L}^H\mathbf{L}^{-H}\mathbf{A}^{-H}\Omega_i^E\mathbf{A}^{-1}\mathbf{L}^{-1}\mathbf{L}\mathbf{A}\mathbf{A}^H\mathbf{L}^H\mathbf{L}^{-H}\mathbf{L}^{-1}\right) \\ &= \frac{1}{\nu-p}\text{Tr}(\mathbf{A}\Omega_i^E\mathbf{A}^H).\end{aligned}\tag{93}$$

Therefore we obtain:

$$\begin{aligned}T_1(i, j) &= \delta_{i,j} \\ T_2(i, j) &= \frac{1}{(\nu-p)^2}\mathbb{E}\left[\text{Tr}(\mathbf{A}\Omega_i^E\mathbf{A}^H)\text{Tr}(\mathbf{A}\Omega_j^E\mathbf{A}^H)\right] \\ T_3(i, j) &= \frac{1}{(\nu-p)^2}\text{Tr}(\Omega_i^E)\mathbb{E}\left[\text{Tr}(\mathbf{A}\Omega_j^E\mathbf{A}^H)\right]\end{aligned}\tag{94}$$

Finally the new formulation for $\mathbf{F}_{\text{prior}}^{AI}$ could be reduced to:

$$\begin{aligned}\left(\mathbf{F}_{\text{prior}}^{AI}\right)_{i,j} &= \left[(\nu+p)^2\text{Tr}(\Omega_i^E)\text{Tr}(\Omega_j^E)\right. \\ &\quad + \mathbb{E}\left[\text{Tr}(\mathbf{A}\Omega_i^E\mathbf{A}^H)\text{Tr}(\mathbf{A}\Omega_j^E\mathbf{A}^H)\right] \\ &\quad - (\nu+p)\text{Tr}(\Omega_i^E)\mathbb{E}\left[\text{Tr}(\mathbf{A}\Omega_j^E\mathbf{A}^H)\right] \\ &\quad \left. - (\nu+p)\text{Tr}(\Omega_j^E)\mathbb{E}\left[\text{Tr}(\mathbf{A}\Omega_i^E\mathbf{A}^H)\right]\right],\end{aligned}\tag{95}$$

where the expectation is taken over Σ . Therefore we have to compute:

$$\begin{aligned}\left(\mathbf{f}_{\text{prior}}^{AI}\right)_i &= \mathbb{E}\left[\text{Tr}(\mathbf{A}\Omega_i^E\mathbf{A}^H)\right] \\ \left(\bar{\mathbf{F}}_{\text{prior}}^{AI}\right)_{i,j} &= \mathbb{E}\left[\text{Tr}(\mathbf{A}\Omega_i^E\mathbf{A}^H)\text{Tr}(\mathbf{A}\Omega_j^E\mathbf{A}^H)\right],\end{aligned}\tag{96}$$

where \mathbf{A} is a lower triangular matrix and where all the elements are independent random variables :

$$\begin{aligned}a_{i,j} &\sim \mathcal{N}_{\mathbb{C}}(0, 1), i > j \\ a_{i,i}^2 &\sim \frac{1}{2}\chi_{2(\nu-i+1)}^2.\end{aligned}\tag{97}$$

C. Computation of (96)

Instead of using the notation Ω_i^E for the basis, we come back to Ω_{ii}^E , Ω_{mn}^E and Ω_{mn}^{h-E} which is recalled in the following:

- Ω_{ii}^E is an p by p symmetric matrix whose i th diagonal element is one, zeros elsewhere
- Ω_{mn}^E is an p by p symmetric matrix whose mn th and nm th elements are both $2^{-1/2}$, zeros elsewhere.

By convention, we use $m > n$.

- Ω_{mn}^{h-E} is an p by p Hermitian matrix whose mn th element is $2^{-1/2}\sqrt{-1}$, and nm th element is $-2^{-1/2}\sqrt{-1}$, zeros elsewhere ($m > n$).

For the calculation of $\mathbf{f}_{\text{prior}}^{AI}$, we have

$$\begin{aligned} \left(\mathbf{f}_{\text{prior}}^{AI}\right)_i &= \mathbb{E}[\text{Tr}(\mathbf{A}\Omega_i^E\mathbf{A}^H)] = \mathbb{E}[\text{Tr}(\Omega_i^E\mathbf{A}^H\mathbf{A})] \\ &= \text{Tr}(\Omega_i^E\mathbb{E}[\mathbf{A}^H\mathbf{A}]). \end{aligned} \quad (98)$$

First, we are looking for the (i, j) –element of the matrix $\mathbf{A}^H\mathbf{A}$. The i^{th} row of the matrix \mathbf{A}^T is the following vector

$$\mathbf{A}^H|_{i^{\text{th}} \text{ row}} = \left(\underbrace{0 \ \cdots \ 0}_{i-1 \text{ elements}} \underbrace{a_{i,i}^* \ a_{i+1,i}^* \ \cdots \ a_{p,i}^*}_{p-i+1 \text{ elements}} \right), \quad (99)$$

and the j^{th} column of the matrix \mathbf{A} is the following vector

$$\mathbf{A}|_{j^{\text{th}} \text{ column}} = \left(\underbrace{0 \ \cdots \ 0}_{j-1 \text{ elements}} \underbrace{a_{j,j} \ a_{j+1,j} \ \cdots \ a_{p,j}}_{p-j+1 \text{ elements}} \right)^T. \quad (100)$$

Then, if $i > j$ the (i, j) –element of the matrix $\mathbf{A}^H\mathbf{A}$ is given by

$$\mathbf{A}^H\mathbf{A}|_{i,j} = \mathbf{A}^H\mathbf{A}|_{j,i} = a_{i,i}^*a_{i,j} + a_{i+1,i}^*a_{i+1,j} + \cdots + a_{p,i}^*a_{p,j}, \quad (101)$$

and $\mathbb{E}[\mathbf{A}^H\mathbf{A}|_{i,j}] = 0$ since all the random variables are independent and $\mathbb{E}[a_{i,j}] = 0$ for $i \neq j$. Consequently, with the same reasoning when $j > i$ one can conclude that $\mathbb{E}[\mathbf{A}^H\mathbf{A}]$ is a diagonal matrix.

If now $i = j$, then

$$\mathbf{A}^H\mathbf{A}|_{i,i} = \underbrace{a_{i,i}^2 + |a_{i+1,i}|^2 + \cdots + |a_{p,i}|^2}_{p-i+1 \text{ elements}}. \quad (102)$$

Since $a_{i,i}^2 \sim \frac{1}{2}\chi_{2(\nu-i+1)}^2$ and $a_{i,j \neq i} \sim \mathcal{CN}(0, 1)$, and the independence of all the random variables, one has $\mathbb{E}[a_{i,i}^2] = \nu - i + 1$ (which appears once) and $\mathbb{E}[|a_{i,j \neq i}|^2] = 1$ (which appears $p - i$ times) and

$$\mathbb{E}[\mathbf{A}^H\mathbf{A}|_{i,i}] = \nu - i + 1 + p - i = \nu + p - 2i + 1. \quad (103)$$

Now if $\Omega_i^E := \Omega_{ii}^E$, one obtains $\left(\mathbf{f}_{\text{prior}}^{AI}\right)_i = \text{Tr}(\Omega_{ii}^E\mathbb{E}[\mathbf{A}^H\mathbf{A}]) = \mathbb{E}[\mathbf{A}^H\mathbf{A}|_{i,i}] = \nu + p - 2i + 1$ and, if $\Omega_i^E := \Omega_{mn}^E$ (or equivalently Ω_{mn}^{h-E}), then $\Omega_{mn}^E\mathbb{E}[\mathbf{A}^H\mathbf{A}]$ is a matrix with zeros on its diagonal and its trace is then equal to zero. Consequently,

$$\left(\mathbf{f}_{\text{prior}}^{AI}\right)_i = \begin{cases} 0 & \text{if } \Omega_i^E := \Omega_{mn}^E \text{ or } \Omega_{mn}^{h-E} \\ \nu + p - 2i + 1 & \text{if } \Omega_i^E := \Omega_{ii}^E \end{cases}. \quad (104)$$

Now we are interested in the calculation of $\bar{\mathbf{F}}_{\text{prior}}^{AI}$. Let's first analyze the term $\text{Tr}(\Omega_i^E\mathbf{A}^H\mathbf{A})$ (since we know from the calculus of $\mathbf{F}_{\text{prior}}^{AI}$ the structure of $\mathbf{A}^H\mathbf{A}|_{i,j}$).

If $\Omega_i^E := \Omega_{ii}^E$ then

$$\text{Tr}(\Omega_{ii}^E \mathbf{A}^H \mathbf{A}) = \mathbf{A}^H \mathbf{A}|_{i,i} = \underbrace{a_{i,i}^2 + |a_{i+1,i}|^2 + \cdots + |a_{p,i}|^2}_{p-i+1 \text{ elements}}. \quad (105)$$

If $\Omega_i^E := \Omega_{mn}^E$ then

$$\begin{aligned} \text{Tr}(\Omega_{mn}^E \mathbf{A}^H \mathbf{A}) &= \frac{1}{\sqrt{2}} (a_{m,m} a_{m,n} + a_{m+1,m}^* a_{m+1,n} + \cdots \\ &\quad + a_{p,m}^* a_{p,n} + a_{m,m} a_{m,n}^* + a_{m+1,m} a_{m+1,n}^* + \cdots + a_{p,m} a_{p,n}^*), \end{aligned} \quad (106)$$

where we have $2(p-i+1)$ elements.

If $\Omega_i^E := \Omega_{mn}^{h-E}$ then

$$\begin{aligned} \text{Tr}(\Omega_{mn}^{h-E} \mathbf{A}^H \mathbf{A}) &= \frac{\sqrt{-1}}{\sqrt{2}} (-a_{m,m} a_{m,n} - a_{m+1,m}^* a_{m+1,n} - \cdots \\ &\quad - a_{p,m}^* a_{p,n} + a_{m,m} a_{m,n}^* + a_{m+1,m} a_{m+1,n}^* + \cdots + a_{p,m} a_{p,n}^*). \end{aligned} \quad (107)$$

Second, let's study $\text{Tr}(\Omega_i^E \mathbf{A}^H \mathbf{A}) \text{Tr}(\Omega_j^E \mathbf{A}^H \mathbf{A})$ and its expectation. Several cases appear:

Case 1: If $\Omega_i^E := \Omega_j^E := \Omega_{ii}^E$

$$\begin{aligned} (\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} &= \mathbb{E}[\text{Tr}^2(\Omega_{ii}^E \mathbf{A}^H \mathbf{A})] \\ &= \mathbb{E} \left[\left(a_{i,i}^2 + |a_{i+1,i}|^2 + \cdots + |a_{p,i}|^2 \right)^2 \right] \\ &= \text{VAR} \left(a_{i,i}^2 + |a_{i+1,i}|^2 + \cdots + |a_{p,i}|^2 \right) \\ &\quad + \left(\mathbb{E} \left[a_{i,i}^2 + |a_{i+1,i}|^2 + \cdots + |a_{p,i}|^2 \right] \right)^2 \\ &= (\nu - i + 1) + (\nu + p - 2i + 1)^2. \end{aligned} \quad (108)$$

Case 2: If $\Omega_i^E := \Omega_{ii}^E$ and $\Omega_j^E := \Omega_{jj}^E$ with $i \neq j$

$$\begin{aligned} (\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} &= \mathbb{E}[\text{Tr}(\Omega_{ii}^E \mathbf{A}^H \mathbf{A}) \text{Tr}(\Omega_{jj}^E \mathbf{A}^H \mathbf{A})] \\ &= \mathbb{E} \left[\left(\underbrace{a_{i,i}^2 + |a_{i+1,i}|^2 + \cdots + |a_{p,i}|^2}_{p-i+1 \text{ elements}} \right) \right. \\ &\quad \left. \left(\underbrace{a_{j,j}^2 + |a_{j+1,j}|^2 + \cdots + |a_{p,j}|^2}_{p-j+1 \text{ elements}} \right) \right] \\ &= (\nu - 2i + p + 1)(\nu - 2j + p + 1). \end{aligned} \quad (109)$$

Case 3: If $\Omega_i^E := \Omega_{ii}^E$ and $\Omega_j^E := \Omega_{mn}^E$ (with $m > n$ but where possibly m or n can be equal to i)

$$\begin{aligned} (\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} &= \frac{1}{\sqrt{2}} \mathbb{E} \left[\left(a_{i,i}^2 + |a_{i+1,i}|^2 + \cdots + |a_{p,i}|^2 \right) \right. \\ &\quad \left(a_{m,m} a_{m,n} + a_{m+1,m}^* a_{m+1,n} + \cdots + a_{p,m}^* a_{p,n} \right. \\ &\quad \left. \left. + a_{m,m} a_{m,n}^* + a_{m+1,m} a_{m+1,n}^* + \cdots + a_{p,m} a_{p,n}^* \right) \right]. \end{aligned} \quad (110)$$

Each terms is the product of an expectation of up to 3 random variables. More precisely, $(\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j}$ is made with terms that can only be in one of following situations: (i) $\mathbb{E}[a_{i,i}^2 a_{m,m} a_{m,n}] = \mathbb{E}[a_{i,i}^2 a_{m,m}] \mathbb{E}[a_{m,n}] = 0$ since $m \neq n$ and $\mathbb{E}[a_{m,n}] = 0$ (same thing when we have $\mathbb{E}[a_{i,i}^2 a_{m,m} a_{m,n}^*]$). (ii) with $N = \{1, \dots, p-m\}$, $\mathbb{E}[a_{i,i}^2 a_{m+N,m}^* a_{m+N,n}] = \mathbb{E}[a_{i,i}^2] \mathbb{E}[a_{m+N,m}^*] \mathbb{E}[a_{m+N,n}] = 0$ since $m \neq n$ and $N > 0$ one always has $a_{m+N,m} \neq a_{m+N,n} \neq a_{i,i}$ (same thing when we have $\mathbb{E}[a_{i,i}^2 a_{m+N,m} a_{m+N,n}^*]$). (iii) with $Q = \{1, \dots, p-m\}$, $\mathbb{E}[a_{i+Q,i}^2 a_{m,m} a_{m,n}] = 0$ because, the random variables $a_{i+Q,i}^2$ and $a_{m,n}$ are the same (i.e. not independent) if and only if $m = i + Q$ and $n = i$ which leads to $\mathbb{E}[a_{i+Q,i}^2 a_{m,m} a_{m,n}] = \mathbb{E}[a_{m,m}] \mathbb{E}[a_{m,n}^3] = 0$ since $m \neq n$ and $\mathbb{E}[a_{m,n}^3] = 0$ (recall that $a_{m,n} \sim \mathcal{CN}(0, 1)$). Otherwise, $a_{i+Q,i}^2$ and $a_{m,n}$ are independent, then $\mathbb{E}[a_{i+Q,i}^2 a_{m,m} a_{m,n}] = \mathbb{E}[a_{m,m}] \mathbb{E}[a_{i+Q,i}^2] \mathbb{E}[a_{m,n}] = 0$ since $m \neq n$ and $\mathbb{E}[a_{m,n}] = 0$. (iv) with $N = \{1, \dots, p-m\}$ and $Q = \{1, \dots, p-m\}$, $\mathbb{E}[|a_{i+Q,i}|^2 a_{m+N,m}^* a_{m+N,n}] = 0$ because $a_{m+N,m} \perp a_{m+N,n}$ since $m > n$ and $N > 0$ (same thing when we have $\mathbb{E}[|a_{i+Q,i}|^2 a_{m+N,m} a_{m+N,n}^*]$). Consequently, for the Case 3, one always has

$$(\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} = 0, \quad (111)$$

and, trivially, the case $\Omega_j^E := \Omega_{jj}^E$ and $\Omega_i^E := \Omega_{mn}^E$ (with $m > n$ but where possibly m or n can be equal to j) leads also to $(\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} = 0$.

Case 4: If $\Omega_i^E := \Omega_{mn}^E$ (with $m > n$) and $\Omega_j^E := \Omega_{kl}^E$ (with $k > l$)

$$\begin{aligned} (\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} &= \frac{1}{2} \mathbb{E} \left[(a_{m,m} a_{m,n} + a_{m+1,n}^* a_{m+1,n} + \dots \right. \\ &\quad \left. + a_{p,m}^* a_{p,n} + a_{m,m} a_{m,n}^* + a_{m+1,m} a_{m+1,n}^* + \dots + a_{p,m} a_{p,n}^*) \right. \\ &\quad \left. (a_{k,k} a_{k,l} + a_{k+1,k}^* a_{k+1,l} + \dots \right. \\ &\quad \left. + a_{p,k}^* a_{p,l} + a_{k,k} a_{k,l}^* + a_{k+1,k} a_{k+1,l}^* + \dots + a_{p,k} a_{p,l}^*) \right]. \end{aligned} \quad (112)$$

Concerning terms such that $\mathbb{E}[a_{m,m} a_{m,n} a_{k,k} a_{k,l}^*]$ (or $\mathbb{E}[a_{m,m} a_{m,n}^* a_{k,k} a_{k,l}]$), if $m = k$ and $n = l$, one has $\mathbb{E}[a_{m,m} a_{m,n} a_{k,k} a_{k,l}^*] = \mathbb{E}[a_{m,m}^2] \mathbb{E}[|a_{m,n}|^2] = \nu - m + 1$ since $a_{m,m}^2 \sim \chi_{\nu-m+1}^2$ and $\mathbb{E}[|a_{m,n}|^2] = 1$. Again if $m = k$ and $n = l$ one also have terms such that $\mathbb{E}[a_{m,m} a_{m,n}^* a_{k,k} a_{k,l}^*] = \mathbb{E}[a_{m,m} a_{m,n} a_{k,k} a_{k,l}] = 0$ since $\mathbb{E}[a_{m,n}^2] = 0$. Otherwise, $\mathbb{E}[a_{m,m} a_{m,n} a_{k,k} a_{k,l}^*] = 0$ and $\mathbb{E}[a_{m,m} a_{m,n}^* a_{k,k} a_{k,l}] = 0$ because $a_{m,n} \perp a_{k,l}$ and $\mathbb{E}[a_{k,l}] = \mathbb{E}[a_{m,n}] = 0$. Concerning terms such that $\mathbb{E}[a_{m,m} a_{m,n} a_{k+N,k}^* a_{k+N,l}]$, $\mathbb{E}[a_{m,m} a_{m,n} a_{k+N,k} a_{k+N,l}^*]$, $\mathbb{E}[a_{m+Q,m}^* a_{m+Q,n} a_{k,k} a_{k,l}]$ and $\mathbb{E}[a_{m+Q,m} a_{m+Q,n}^* a_{k,k} a_{k,l}]$, they are always equal to 0 since $a_{k+N,k} \perp a_{k+N,l}$ and $a_{m+Q,m} \perp a_{m+Q,n}$.

Finally, $\mathbb{E}[a_{m+Q,m}^* a_{m+Q,n} a_{k+N,k}^* a_{k+N,l}] = \mathbb{E}[a_{m+Q,m} a_{m+Q,n}^* a_{k+N,k} a_{k+N,l}] = \mathbb{E}[|a_{m+Q,m}|^2 |a_{m+Q,n}|^2] = 1$ if and only if $m = k$, $n = l$ and $P = N$ (otherwise one always have at least on zero mean random

variable independent other the three others, or the fact that $\mathbb{E}[a_{m,n}^2] = 0$, and these terms are equal to 0). Consequently for case 4, if $\Omega_i^E := \Omega_{mn}^E = \Omega_j^E$

$$(\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} = (\nu - m + 1 + p - m) = (\nu - 2m + p + 1), \quad (113)$$

and $(\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} = 0$ otherwise.

Case 5: If $\Omega_i^E := \Omega_{ii}^E$ and $\Omega_j^E := \Omega_{mn}^{h-E}$ (with $m > n$ but where possibly m or n can be equal to i)

$$\begin{aligned} (\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} &= \frac{\sqrt{-1}}{\sqrt{2}} \mathbb{E} \left[\left(a_{i,i}^2 + |a_{i+1,i}|^2 + \cdots + |a_{p,i}|^2 \right) \right. \\ &\quad \left(-a_{m,m}a_{m,n} - a_{m+1,m}^*a_{m+1,n} - \cdots - a_{p,m}^*a_{p,n} \right. \\ &\quad \left. \left. + a_{m,m}a_{m,n}^* + a_{m+1,m}a_{m+1,n}^* + \cdots + a_{p,m}a_{p,n}^* \right) \right]. \end{aligned} \quad (114)$$

An analysis, similar to the one provided in case 3, and using the fact that $\forall N = \{1, \dots, p-m\}$, $\mathbb{E} \left[a_{m+N,m}^*a_{m+N,n}a_{m+N,m}^* \right] = \mathbb{E} \left[a_{m+N,m}^*a_{m+N,n}a_{m+N,n} \right] = 0$ leads to

$$(\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} = 0. \quad (115)$$

Of course, the case $\Omega_j^E := \Omega_{jj}^E$ and $\Omega_i^E := \Omega_{mn}^{h-E}$ (with $m > n$ but where possibly m or n can be equal to j) leads also to $(\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} = 0$.

Case 6: If $\Omega_i^E := \Omega_{mn}^E$ (with $m > n$) and $\Omega_j^E := \Omega_{kl}^{h-E}$ (with $k > l$)

$$\begin{aligned} (\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} &= \frac{\sqrt{-1}}{2} \mathbb{E} \left[\left(a_{m,m}a_{m,n} + a_{m+1,m}^*a_{m+1,n} + \cdots \right. \right. \\ &\quad \left. \left. + a_{p,m}^*a_{p,n} + a_{m,m}a_{m,n}^* + a_{m+1,m}a_{m+1,n}^* + \cdots + a_{p,m}a_{p,n}^* \right) \right. \\ &\quad \left(-a_{k,k}a_{k,l} - a_{k+1,k}^*a_{k+1,l} - \cdots \right. \\ &\quad \left. \left. - a_{p,k}^*a_{p,l} + a_{k,k}a_{k,l}^* + a_{k+1,k}a_{k+1,l}^* + \cdots + a_{p,k}a_{p,l}^* \right) \right]. \end{aligned} \quad (116)$$

Using all the aforementioned arguments, one find after calculus that

$$(\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} = 0. \quad (117)$$

Of course, the case $\Omega_i^E := \Omega_{mn}^{h-E}$ (with $m > n$) and $\Omega_j^E := \Omega_{kl}^E$ (with $k > l$) leads also to $(\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} = 0$.

Case 7: If $\Omega_i^E := \Omega_{mn}^{h-E}$ (with $m > n$) and $\Omega_j^E := \Omega_{kl}^{h-E}$ (with $k > l$)

$$\begin{aligned} (\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} &= -\frac{1}{2} \mathbb{E} \left[\left(-a_{m,m}a_{m,n} - a_{m+1,m}^*a_{m+1,n} - \cdots \right. \right. \\ &\quad \left. \left. - a_{p,m}^*a_{p,n} + a_{m,m}a_{m,n}^* + a_{m+1,m}a_{m+1,n}^* + \cdots + a_{p,m}a_{p,n}^* \right) \right. \\ &\quad \left(-a_{k,k}a_{k,l} - a_{k+1,k}^*a_{k+1,l} - \cdots \right. \\ &\quad \left. \left. - a_{p,k}^*a_{p,l} + a_{k,k}a_{k,l}^* + a_{k+1,k}a_{k+1,l}^* + \cdots + a_{p,k}a_{p,l}^* \right) \right]. \end{aligned} \quad (118)$$

An analysis, similar to the one provided in case 4, shows that $(\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} \neq 0$ if and only if $m = k$ and $n = l$. In this case, one obtains easily

$$(\bar{\mathbf{F}}_{\text{prior}}^{AI})_{i,j} = \nu - 2m + p + 1. \quad (119)$$

Final result: Then one can summary the value of $\left(\bar{\mathbf{F}}_{\text{prior}}^{AI}\right)_{i,j} \neq 0$ as:

- If $\Omega_i^E := \Omega_{ii}^E := \Omega_j^E$ then $\left(\bar{\mathbf{F}}_{\text{prior}}^{AI}\right)_{i,j} = (\nu - i + 1) + (\nu + p - 2i + 1)^2$.
- If $\Omega_i^E := \Omega_{ii}^E$ and $\Omega_j^E := \Omega_{jj}^E$ with $i \neq j$ then $\left(\bar{\mathbf{F}}_{\text{prior}}^{AI}\right)_{i,j} = (\nu - 2i + p + 1)(\nu - 2j + p + 1)$.
- If $\Omega_i^E := \Omega_{mn}^E = \Omega_j^E$ then $\left(\bar{\mathbf{F}}_{\text{prior}}^{AI}\right)_{i,j} = \nu - 2m + p + 1$.
- If $\Omega_i^E := \Omega_{mn}^{h-E} = \Omega_j^E$ then $\left(\bar{\mathbf{F}}_{\text{prior}}^{AI}\right)_{i,j} = \nu - 2m + p + 1$.