RIEMANNIAN GEOMETRY AND CRAMÉR-RAO BOUND FOR BLIND SEPARATION OF GAUSSIAN SOURCES

Florent Bouchard[⋆] *Arnaud Breloy*[†] *Alexandre Renaux*[‡] *Guillaume Ginolhac*[⋆]

[⋆] LISTIC, Univ. Savoie Mont Blanc, Annecy, France
[†] LEME, Univ. Paris Nanterre, France
[‡] L2S, Univ. Paris-Sud,, France

ABSTRACT

We consider the optimal performance of blind separation of Gaussian sources. In practice, this estimation problem is solved by a two-step procedure: estimation of a set of covariance matrices from the observed data and approximate joint diagonalization of this set to find the unmixing matrix. Rather than studying the theoretical performance of a specific method, we are interested in the optimal attainable performance of any estimator. To do so, we consider the so-called intrinsic Cramér-Rao bound, which exploits the geometry of the parameters of the model. Unlike previous works developing a Cramér-Rao bound in this context, our solution does not require any additional hypotheses. To obtain our bound, we define and study a new Riemannian manifold holding the parameters of interest. An original estimation error measure is defined with the help of our Riemannian distance function. The corresponding Fisher information matrix is then obtained from the Fisher information metric and orthonormal bases on the tangent spaces of the manifold. Finally, our theoretical results are validated on simulated data.

***Index Terms*—** intrinsic Cramér-Rao bound, blind source separation, approximate joint diagonalization, Riemannian geometry

1. INTRODUCTION

Blind source separation is a major tool for signal processing and data analysis in a wide range of engineering fields such as radar, communications, image processing and biomedical signals analysis; see [1] for a review of theory and applications. We consider the determined linear instantaneous blind separation of Gaussian sources problem based on the mixing model

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}, \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ corresponds to the observations, $\boldsymbol{s} \in \mathbb{R}^n$ is the centered multivariate Gaussian random variable with independent components corresponding to the sources and $\boldsymbol{A} \in \mathrm{GL}_n$ ($n \times n$ non-singular matrices) is the mixing matrix. Given some observations of $\boldsymbol{x}$, the aim is to retrieve estimates $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{s}})$ of $(\boldsymbol{A}, \boldsymbol{s})$.

In practice, we have K sets $\{\boldsymbol{x}_k(t)\}$ of T observations. Since the sources are independent, the covariance matrix of $\{\boldsymbol{s}_k(t)\}$ is $\boldsymbol{\Lambda}_k \in \mathcal{D}_n^{++}$ (diagonal positive definite matrices). Thus, $\{\boldsymbol{x}_k(t)\}$ follows the multivariate Gaussian distribution with covariance matrix

$$\boldsymbol{C}_k = \boldsymbol{A}\boldsymbol{\Lambda}_k\boldsymbol{A}^T \in \mathcal{S}_n^{++} \tag{2}$$

(symmetric positive definite matrices) and the estimation problem consists in retrieving $(\boldsymbol{A}, \{\boldsymbol{\Lambda}_k\})$. To solve it, most methods follow the two steps procedure: *(i)* compute estimators $\widehat{\boldsymbol{C}}_k$ of the K covariance matrices of $\{\boldsymbol{x}_k(t)\}$; *(ii)* perform approximate joint diagonalization of $\{\widehat{\boldsymbol{C}}_k\}$, *i.e.*, find $\boldsymbol{B} \in \mathrm{GL}_n$ such that $\{\boldsymbol{B}\widehat{\boldsymbol{C}}_k\boldsymbol{B}^T\}$ contains matrices as diagonal as possible according to a diagonality criterion. The estimator $(\widehat{\boldsymbol{A}}, \{\widehat{\boldsymbol{\Lambda}}_k\})$ is then defined as $(\boldsymbol{B}^{-1}, \boldsymbol{B}\widehat{\boldsymbol{C}}_k\boldsymbol{B}^T)$.

The two most popular diagonality criteria are: *(i)* the least squares criterion, proposed in [2], which is based on the Frobenius distance between $\boldsymbol{B}\widehat{\boldsymbol{C}}_k\boldsymbol{B}^T$ and its diagonal part; and *(ii)* the criterion based on the log-likelihood of model (1) [3,4]. Other diagonality criteria have also been considered; see *e.g.* [5, 6] which exploits the geometry of $\mathcal{S}_n^{++}$. Many methods have been developed with various criteria; see *e.g.* [2,4,7–9]. Recently, the geometrical structure of GL_n has been exploited to build a Riemannian optimization framework adapted to approximate joint diagonalization [6, 10].

Concerning the theoretical performance of the estimators, different studies exist for the considered model. An asymptotic analysis is given in [3] for the log-likelihood criterion. A theoretical analysis in terms of interference signal ratio is proposed in [11] for the least squares criterion. More generally, several papers propose a theoretical analysis of their estimator, as in [12, 13] for a more general data model. In [8], the author derives the theoretical analysis of algorithms based either on the least squares or log-likelihood criteria when model errors occur. However, all these analyses do not answer the question of the optimal attainable performance for the con-

This work was supported by MARGARITA ANR-17-ASTR-0015 and PHOENIX ANR-15-CE23-0012 grants of the French National Agency of Research.

sidered model. To achieve this, one can use the inequality between the mean squared error (MSE) and the Cramér-Rao bound (CRB)

$$\mathrm{MSE}(\theta, \widehat{\theta}) \geq \mathrm{CRB}(\theta),$$

where θ contains the true parameters and $\widehat{\theta}$ is an estimator. In the context of the model that we consider, [14] derives the Cramér-Rao bound associated with the interference signal ratio for the estimation problem of $\{\boldsymbol{C}_k\}$ and $\boldsymbol{A}$ (equivalent to our problem of estimating $\boldsymbol{A}$ and $\{\boldsymbol{\Lambda}_k\}$). To obtain the inequality in this case, it is needed for the blocks between parameters $\{\boldsymbol{C}_k\}$ and $\boldsymbol{A}$ in the Fisher information matrix to be equal to $\boldsymbol{0}$. Unfortunatly, as mentionned in [14], this property does not hold in general within our setting.

To overcome this issue, we propose to derive a new inequality by using an original error measure and the corresponding so-called intrinsic Cramér-Rao bound theoretically defined in [15] which fits well with the source separation problem. Inspired by the works of [16] and using the procedure in [17], the estimation error is measured by a new Riemannian distance. The associated Fisher information matrix is obtained from the Fisher information metric and orthonormal bases of the tangent spaces of the parameter manifold.

The geometry of the manifold $\mathrm{GL}_n \times (\mathcal{D}_n^{++})^K$ holding the parameters follows from those of GL_n and $\mathcal{D}_n^{++}$. Since sources are Gaussian, we consider the well known geometry of $\mathcal{D}_n^{++}$ corresponding to the Gaussian distribution [18]. Concerning GL_n, several geometries have been considered. Treating it directly, as in *e.g.* [10, 19, 20], appears complicated in our case since the Riemannian distance function is not known in closed form. We rather consider the geometry proposed in [6], where the polar decomposition is exploited to define an isomorphic manifold to GL_n: the product $\mathcal{O}_n \times \mathcal{S}_n^{++}$ ($\mathcal{O}_n$, orthogonal matrices). This representation appears advantageous because Riemannian distances for both $\mathcal{O}_n$ and $\mathcal{S}_n^{++}$ are known.

2. MODEL

In this section, we detail the distribution of the data $\{\boldsymbol{x}_k^{(t)}\}$ at hand and we show that the parameters $(\boldsymbol{A}, \{\boldsymbol{\Lambda}_k\})$ belong to the manifold $\mathcal{M} = (\mathcal{O}_n \times \mathcal{S}_n^{++})$. Then, we study the geometry of $\mathcal{M}$ in order to define the Riemannian distance function and the corresponding orthonormal bases on the tangent spaces of $\mathcal{M}$, which are needed to define the error measure and Fisher information matrix, respectively.

2.1. Data and parameter manifold

The K sets of observations $\{\boldsymbol{x}_k^{(t)}\}$ in $\mathbb{R}^n$ follow the centered mutivariate Gaussian distribution with covariance matrices $\{\boldsymbol{A}\boldsymbol{\Lambda}_k\boldsymbol{A}^T\} \in \mathcal{S}_n^{++}$, where $\boldsymbol{A} \in \mathrm{GL}_n$ and $\boldsymbol{\Lambda}_k \in \mathcal{D}_n^{++}$. Thus, the probability density function f of $\{\boldsymbol{x}_k^{(t)}\}$ is

$$f(\{\boldsymbol{x}_k(t)\}|\boldsymbol{A}, \{\boldsymbol{\Lambda}_k\}) = \prod_k f_{\mathrm{G}}(\{\boldsymbol{x}_k(t)\}|\boldsymbol{A}\boldsymbol{\Lambda}_k\boldsymbol{A}^T), \quad (3)$$

where f_{G} is the probability density function of the centered multivariate Gaussian distribution. Given $\{\boldsymbol{x}(t)\}$ with covariance matrix $\boldsymbol{C}$, we have, up to a factor,

$$f_{\mathrm{G}}(\{\boldsymbol{x}(t)\}|\boldsymbol{C}) = \prod_t \det(\boldsymbol{C})^{-1/2} \exp(-\boldsymbol{x}(t)^T\boldsymbol{C}^{-1}\boldsymbol{x}(t)/2). \quad (4)$$

The set of parameters $(\boldsymbol{A}, \boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_K)$ of the distribution of $\{\boldsymbol{x}_k\}$ lie in the manifold $\mathrm{GL}_n \times (\mathcal{D}_n^{++})^K$ (it is a manifold since it is the product of $K+1$ manifolds). As explained in the introduction, dealing with GL_n directly is complicated in our case since the Riemannian distance is not known in closed form. Instead, we exploit the polar decomposition as in [6]: every matrix $\boldsymbol{A} \in \mathrm{GL}_n$ admits the unique decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}$, where $(\boldsymbol{U}, \boldsymbol{S}) \in \mathcal{O}_n \times \mathcal{S}_n^{++}$. Thus, $\mathcal{O}_n \times \mathcal{S}_n^{++}$ is isomorphic to GL_n and we can choose $\mathcal{M} = (\mathcal{O}_n \times \mathcal{S}_n^{++}) \times (\mathcal{D}_n^{++})^K$ as the parameter manifold.

2.2. Riemannian geometry of $\mathcal{M}$

In the following, $\theta = (\boldsymbol{U}, \boldsymbol{S}, \{\boldsymbol{\Lambda}_k\})$, $\xi = (\boldsymbol{\xi}_{\boldsymbol{U}}, \boldsymbol{\xi}_{\boldsymbol{S}}, \{\boldsymbol{\xi}_k\})$ and $\eta = (\boldsymbol{\eta}_{\boldsymbol{U}}, \boldsymbol{\eta}_{\boldsymbol{S}}, \{\boldsymbol{\eta}_k\})$. The tangent space $T_\theta\mathcal{M}$ of $\theta \in \mathcal{M}$ is

$$T_\theta\mathcal{M} = \{\xi \in \mathbb{R}^{n\times n} \times \mathcal{S}_n \times (\mathcal{D}_n)^K : \boldsymbol{\xi}_{\boldsymbol{U}} = \boldsymbol{U}\boldsymbol{\Omega}, \boldsymbol{\Omega}^T = -\boldsymbol{\Omega}\},$$

where $\mathcal{S}_n$ and $\mathcal{D}_n$ denote the spaces of symmetric and diagonal matrices, respectively. We equip $\mathcal{M}$ with the Riemannian metric

$$\langle \xi, \eta \rangle_\theta = \mathrm{tr}(\boldsymbol{\xi}_{\boldsymbol{U}}^T\boldsymbol{\eta}_{\boldsymbol{U}}) + \mathrm{tr}(\boldsymbol{S}^{-1}\boldsymbol{\xi}_{\boldsymbol{S}}\boldsymbol{S}^{-1}\boldsymbol{\eta}_{\boldsymbol{S}}) + \sum_k \mathrm{tr}(\boldsymbol{\Lambda}_k^{-1}\boldsymbol{\xi}_k\boldsymbol{\Lambda}_k^{-1}\boldsymbol{\eta}_k). \quad (5)$$

The Riemannian distance function on $\mathcal{M}$ resulting from metric (5), which is crucial to define the error between true parameters θ and estimator $\widehat{\theta}$, is given in proposition 1.

Proposition 1. *The Riemannian distance on $\mathcal{M}$ resulting from* (5) *is, given $\theta = (\boldsymbol{U}, \boldsymbol{S}, \{\boldsymbol{\Lambda}_k\})$ and $\widehat{\theta} = (\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{S}}, \{\widehat{\boldsymbol{\Lambda}}_k\})$,*

$$\delta^2(\theta, \widehat{\theta}) = \left\|\log(\boldsymbol{U}^T\widehat{\boldsymbol{U}})\right\|_F^2 + \left\|\log(\boldsymbol{S}^{-1/2}\widehat{\boldsymbol{S}}\boldsymbol{S}^{-1/2})\right\|_F^2 + \sum_k \left\|\log(\boldsymbol{\Lambda}_k^{-1}\widehat{\boldsymbol{\Lambda}}_k)\right\|_F^2,$$

where log *denotes the matrix logarithm.*

Proof. The Riemannian distance on $\mathcal{O}_n$ associated with the part of the metric (5) that concerns $\boldsymbol{U}$ is $\|\log(\boldsymbol{U}^T\widehat{\boldsymbol{U}})\|_F$ [21, 22]. Simarly, the distances on $\mathcal{S}_n^{++}$ and $\mathcal{D}_n^{++}$ associated with the parts that concern $\boldsymbol{S}$ and $\boldsymbol{\Lambda}_k$ are $\|\log(\boldsymbol{S}^{-1/2}\widehat{\boldsymbol{S}}\boldsymbol{S}^{-1/2})\|_F$ and $\|\log(\boldsymbol{\Lambda}_k^{-1}\widehat{\boldsymbol{\Lambda}}_k)\|_F$ [15, 18]. The result then follows from the properties of product manifolds. □

Finally, an orthonormal basis in $T_\theta\mathcal{M}$ according to metric (5), which is needed to define the Fisher information matrix associated with the estimation problem on $\mathcal{M}$, is given in proposition 2.

Proposition 2. *An orthonormal basis $\{e_\theta^p\}_p$ in $T_\theta\mathcal{M}$ according to metric* (5) *is defined, for $i, j \in \{1, \dots, n\}$, by*

$$\Big\{\{(\boldsymbol{e}_{\boldsymbol{U}}^{ij}, \mathbf{0}, \mathbf{0}, \dots, \mathbf{0})\}_{i>j}, \{(\mathbf{0}, \boldsymbol{e}_{\boldsymbol{S}}^{ij}, \mathbf{0}, \dots, \mathbf{0})\}_{i\geq j}, \\ \{(\mathbf{0}, \mathbf{0}, \boldsymbol{e}_1^i, \mathbf{0}, \dots, \mathbf{0})\}_i, \{(\mathbf{0}, \mathbf{0}, \mathbf{0}, \dots, \mathbf{0}, \boldsymbol{e}_K^i)\}_i\Big\},$$

where

- $\boldsymbol{e}_{\boldsymbol{U}}^{ij} = \boldsymbol{U}\boldsymbol{\Omega}^{ij}$*:* $\boldsymbol{\Omega}^{ij}$ *skew-symmetric matrix whose* ij^{th} *and* ji^{th} *elements are* $1/\sqrt{2}$ *and* $-1/\sqrt{2}$*, zeros elsewhere.*
- $\boldsymbol{e}_{\boldsymbol{S}}^{ij} = \boldsymbol{S}^{1/2}\boldsymbol{\xi}_{\boldsymbol{S}}^{ij}\boldsymbol{S}^{1/2}$*:* $\boldsymbol{\xi}_{\boldsymbol{S}}^{ii}$ *diagonal matrix whose* i^{th} *diagonal element is one, zeros elsewhere;* $\boldsymbol{\xi}_{\boldsymbol{S}}^{ij}$*,* $i > j$*, symmetric matrix whose* ij^{th} *and* ji^{th} *elements are* $1/\sqrt{2}$*, zeros elsewhere.*
- $\boldsymbol{e}_k^i = \boldsymbol{\Lambda}_k^{1/2}\boldsymbol{\xi}_{\boldsymbol{\Lambda}}^i\boldsymbol{\Lambda}_k^{1/2}$*:* $\boldsymbol{\xi}_{\boldsymbol{\Lambda}}^i$ *diagonal matrix whose* i^{th} *element is equal to one, zeros elsewhere.*

Proof. By definition, it suffices to check that $\langle e_\theta^p, e_\theta^p\rangle_\theta = 1$ and $\langle e_\theta^p, e_\theta^q\rangle_\theta = 0$, $p \neq q$. Basic calculations yield the result. □

3. INTRINSIC CRAMÉR-RAO BOUND

We now have all the geometrical elements required to define the intrinsic Cramér-Rao bound for the considered estimation problem on $\mathcal{M}$. In order to do so, we first define our estimation error measure from the Riemannian distance of proposition 1. Then, we obtain the Fisher information metric of distribution (3) on $\mathcal{M}$ and we construct the Fisher information matrix associated to the basis of proposition 2. Finally, we obtain the wished inequality.

3.1. Estimation error measure

We define how the error of an unbiased estimator $\widehat{\theta}$ of θ in $\mathcal{M}$ is measured. From [15, 16], we know that it corresponds to the squared Riemannian distance δ^2 on $\mathcal{M}$ between θ and $\widehat{\theta}$. It is well known that an estimator $\widehat{\theta}$ of the blind source separation problem is only defined up to a permutation and diagonal scaling. Given any $\boldsymbol{P} \in \mathbb{P}_n$ (permutation matrices) and $\boldsymbol{\Sigma} \in \mathcal{D}_n^*$ (non-singular diagonal matrices), $\widehat{\theta}$ is equivalent to

$$\widehat{\theta}(\boldsymbol{P}, \boldsymbol{\Sigma}) = \Big(\widehat{\boldsymbol{U}}(\boldsymbol{P}, \boldsymbol{\Sigma}), \widehat{\boldsymbol{S}}(\boldsymbol{P}, \boldsymbol{\Sigma}), \\ \boldsymbol{P}^T\boldsymbol{\Sigma}^{-2}\widehat{\boldsymbol{\Lambda}}_1\boldsymbol{P}, \dots, \boldsymbol{P}^T\boldsymbol{\Sigma}^{-2}\widehat{\boldsymbol{\Lambda}}_K\boldsymbol{P}\Big),$$

where $(\widehat{\boldsymbol{U}}(\boldsymbol{P}, \boldsymbol{\Sigma}), \widehat{\boldsymbol{S}}(\boldsymbol{P}, \boldsymbol{\Sigma})) \in \mathcal{O}_n \times \mathcal{S}_n^{++}$ corresponds to the polar decomposition of $\widehat{\boldsymbol{U}}\widehat{\boldsymbol{S}}\boldsymbol{\Sigma}\boldsymbol{P}$.

Let $\widehat{\theta}_*$ be the estimator equivalent to $\widehat{\theta}$ that best corresponds to θ in $\mathcal{M}$, *i.e.*, such that

$$\widehat{\theta}_* = \min_{\boldsymbol{P}\in\mathbb{P}_n, \boldsymbol{\Sigma}\in\mathcal{D}_n^*} \delta_{\mathcal{M}}^2(\theta, \widehat{\theta}(\boldsymbol{P}, \boldsymbol{\Sigma})).$$

It follows that the error of the unbiased estimator $\widehat{\theta}$ of θ is defined as $\delta_{\mathcal{M}}^2(\theta, \widehat{\theta}_*)$.

3.2. Fisher information matrix

To obtain the Fisher information matrix, we first need to derive the Fisher information metric on $\mathcal{M}$ associated with the probability density function (3) of our estimation problem. It is achieved in proposition 3.

Proposition 3. *Let $\theta = \in \mathcal{M}$, $\xi, \eta \in T_\theta\mathcal{M}$ and the mapping φ_k from $\mathcal{M}$ onto $\mathcal{S}_n^{++}$ defined by $\varphi_k(\theta) = \boldsymbol{U}\boldsymbol{S}\boldsymbol{\Lambda}_k\boldsymbol{S}\boldsymbol{U}^T$. The Fisher information metric g_θ^{FIM} on $\mathcal{M}$ of the probability density function* (3) *is*

$$g_\theta^{\text{FIM}}(\xi, \eta) = \sum_k g^{\text{FIM}}_{\text{G}\,\varphi_k(\theta)}(\text{D}\,\varphi_k(\theta)[\xi], \text{D}\,\varphi_k(\theta)[\eta]),$$

where

$$g^{\text{FIM}}_{\text{G}\,\boldsymbol{C}}(\boldsymbol{\xi}, \boldsymbol{\eta}) = \frac{T}{2}\,\text{tr}(\boldsymbol{C}^{-1}\boldsymbol{\xi}\boldsymbol{C}^{-1}\boldsymbol{\eta})$$

is the well-known Gaussian Fisher information metric [15, 18], defined for all $\boldsymbol{C} \in \mathcal{S}_n^{++}$ and $\boldsymbol{\xi}$, $\boldsymbol{\eta} \in \mathcal{S}_n$; and the directional derivative of φ_k is

$$\text{D}\,\varphi_k(\theta)[\xi] = (\boldsymbol{U}\boldsymbol{\xi}_{\boldsymbol{S}} + \boldsymbol{\xi}_{\boldsymbol{U}}\boldsymbol{S})\boldsymbol{\Lambda}_k\boldsymbol{S}\boldsymbol{U}^T + \boldsymbol{U}\boldsymbol{S}\boldsymbol{\xi}_k\boldsymbol{S}\boldsymbol{U}^T \\ + \boldsymbol{U}\boldsymbol{S}\boldsymbol{\Lambda}_k(\boldsymbol{S}\boldsymbol{\xi}_{\boldsymbol{U}}^T + \boldsymbol{\xi}_{\boldsymbol{S}}\boldsymbol{U}^T).$$

Proof. One can check that the log-likelihood L of (3) is

$$L(\theta) = \sum_k L_{\text{G}} \circ \varphi_k(\theta),$$

where L_{G} is the log-likelihood of the Gaussian distribution (4). By definition, $g_\theta^{\text{FIM}}(\xi, \eta) = \mathbb{E}[\text{D}\,L(\theta)[\xi]\,\text{D}\,L(\theta)[\eta]]$. From [15, theorem 1], we have

$$\begin{aligned} g_\theta^{\text{FIM}}(\xi, \eta) &= -\mathbb{E}[\text{D}^2\,L(\theta)[\xi, \eta]] \\ &= -\textstyle\sum_k \mathbb{E}[\text{D}^2(L_{\text{G}} \circ \varphi_k)(\theta)[\xi, \eta]] \\ &= \textstyle\sum_k \mathbb{E}[\text{D}(L_{\text{G}} \circ \varphi_k)(\theta)[\xi]\,\text{D}(L_{\text{G}} \circ \varphi_k)(\theta)[\eta] \\ &= \textstyle\sum_k g^{\text{FIM}}_{\text{G}\,\varphi_k(\theta)}(\text{D}\,\varphi_k(\theta)[\xi], \text{D}\,\varphi_k(\theta)[\eta]). \end{aligned}$$

□

Notice that g_θ^{FIM} does not define a Riemannian metric on $\mathcal{M}$ in this case as it is not positive definite. This is due to the invariance with respect to the action of $\boldsymbol{\Sigma} \in \mathcal{D}_n^*$: θ and $\widetilde{\theta}$, such that $\boldsymbol{U}\boldsymbol{S} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{S}}\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}_k = \boldsymbol{\Sigma}^{-2}\widetilde{\boldsymbol{\Lambda}}_k$, are equivalent and $\varphi_k(\theta) = \varphi_k(\widetilde{\theta})$. As φ_k is constant on the fiber associated with $\boldsymbol{\Sigma}$, its directional derivative vanishes.

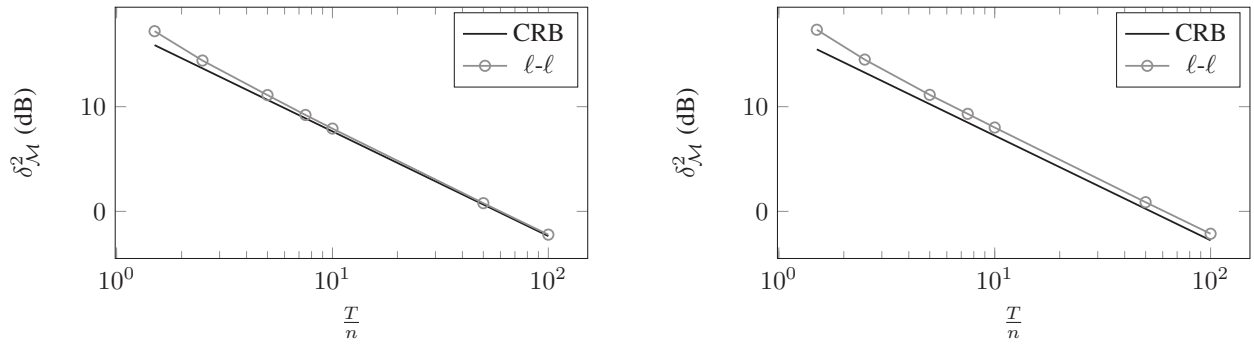

Fig. 1. Intrinsic Cramér-Rao-Rao bound (CRB) and mean over 100 trials of the squared distance of approximate joint diagonalization estimator (ℓ-ℓ) based on the log-likelihood criterion minimized with the framework built in [6] with the intrinsic constraint versus T/n. The condition number with respect to inversion α of $\boldsymbol{A}$ is 10 on the left panel and 1000 on the right panel. In this simulation, $K = 30$ and $n = 10$.

From the Fisher information metric of proposition 3 and the orthonormal basis $\{\xi^p_\theta\}_p$ of $T_\theta\mathcal{M}$ defined in proposition 2, we can construct the Fisher information matrix $\boldsymbol{F}$ of size $(n^2 + nK) \times (n^2 + nK)$. Its pq^{th} element is defined as

$$\boldsymbol{F}_{pq} = g^{\text{FIM}}_\theta(\xi^p_\theta, \xi^q_\theta).$$

Because of the invariance with respect to the action of non-singular diagonal matrices described above, the Fisher information matrix is singular and its rank is $n^2 + nK - n$.

3.3. Inequality

We finally derive the intrinsic Cramér-Rao bound [15, 16] of an unbiased estimator $\widehat{\theta}$ of θ in $\mathcal{M}$. Our Fisher information matrix $\boldsymbol{F}$ is singular. In such a case, the inverse of $\boldsymbol{F}$ can be replaced by the Moore-Penrose pseudo-inverse in the bound as explained in [16]. It follows that the intrinsic Cramér-Rao bound is

$$\mathbb{E}\left[\delta^2_{\mathcal{M}}(\theta, \widehat{\theta}_*)\right] \geq \text{tr}(\boldsymbol{F}^\dagger), \tag{6}$$

where $\cdot^\dagger$ denotes the Moore-Penrose pseudo-inverse.

4. NUMERICAL ILLUSTRATION

We generate a set of $K = 30$ $n \times n$ (with $n = 10$) symmetric positive definite matrices $\boldsymbol{C}_k$ according to model

$$\boldsymbol{C}_k = \boldsymbol{A}\boldsymbol{\Lambda}_k\boldsymbol{A}^T. \tag{7}$$

$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, with $\boldsymbol{U}$ and $\boldsymbol{V}$ random orthogonal matrices, and $\boldsymbol{\Sigma}$ random diagonal matrix whose minimal and maximal elements are $1/\sqrt{\alpha}$ and $\sqrt{\alpha}$, where $\alpha \in \{10, 1000\}$ is the condition number of $\boldsymbol{A}$ with respect to inversion. Matrices $\boldsymbol{\Lambda}_k \in \mathcal{D}^{++}_n$ hold source energies and have i.i.d elements drawn from the chi-squared distribution with expectation 1. For $T \in \{15, 50, 100, 500, 1000\}$, we compute 100 sets of T random realizations $\{\boldsymbol{x}_k(t)\}$ drawn from the centered multivariate Gaussian distribution with covariance matrices $\boldsymbol{C}_k$. Notice that, with these simulated data, as the determinants of matrices $\boldsymbol{C}_k$ are not assumed to be known, the hypothesis in [14] for the bound inequality to be derived is not verified. Moreover, the blocks between $\boldsymbol{A}$ and $\boldsymbol{\Lambda}_k$ in our proposed Fisher information matrix are not equal to $\boldsymbol{0}$ in general with this model.

To perform the blind source separation of $\{\boldsymbol{x}_k(t)\}$, we first estimate the sample covariance matrices $\widehat{\boldsymbol{C}}_k = T^{-1}\sum_t \boldsymbol{x}_k(t)\boldsymbol{x}_k(t)^T$. Then, we compute the joint diagonalizer $\boldsymbol{B}$ of the set $\{\boldsymbol{C}_k\}$ by employing the approximate joint diagonalization framework proposed in [6] with the criterion based on the log-likelihood [3, 4]. We choose the so-called intrinsic constraint in [6], which imposes a specific scaling of the rows of $\boldsymbol{B}$, in order to avoid diagonal scaling issues.

In figure 1, the intrinsic Cramér-Rao bound inequality (6) is illustrated for the joint diagonalization estimator based on the log-likelihood for different values of α and T/n. For $\alpha = 10$, we observe that the performance of this estimator gets close to the intrinsic Cramér-Rao bound as T/n grows. For $\alpha = 1000$, the performance gets closer to the bound as T/n grows, without reaching it. Thus, this simulation suggests that our intrinsic Cramér-Rao bound is a good predictor of attainable performance and that a more accurate estimator might be found for illed-conditioned mixing matrices.

5. CONCLUSIONS

In this article, we have proposed an intrinsic Cramér-Rao bound for blind separation of Gaussian sources, which is solved through approximate joint diagonalization of a set of estimated covariance matrices of the data. We have defined the Riemannian geometry of a new manifold which holds the parameters of the model. We have derived the Fisher information metric associated with the model and the corresponding bound. We conclude that this intrinsic Cramér-Rao bound is a good predictor of performance, thus an interesting alternative to theoretical performance studies of specific estimators.

6. REFERENCES

[1] P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, 1st edition, 2010.

[2] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEEE Proceedings-F*, 140(6):362–370, dec 1993.

[3] D. T. Pham and P. Garat. Blind beamforming for non Gaussian signals. *IEEE Transactions on Signal Processing*, 45(7):1712–1725, July 1997.

[4] D. T. Pham and J.-F. Cardoso. Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Transactions on Signal Processing*, 49(9):1837–1848, 2001.

[5] K. Alyani, M. Congedo, and M. Moakher. Diagonality measures of Hermitian positive-definite matrices with application to the approximate joint diagonalization problem. *Linear Algebra and its Applications*, 2016.

[6] F. Bouchard, J. Malick, and M. Congedo. Riemannian optimization and approximate joint diagonalization for blind source separation. *IEEE Transactions on Signal Processing*, 66(8):2041–2054, 2018.

[7] A. Ziehe, P. Laskov, G. Nolte, and K.-R. Müller. A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. *The Journal of Machine Learning Research*, 5:777–800, 2004.

[8] B. Afsari. Sensitivity analysis for the problem of matrix joint diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1148–1171, 2008.

[9] P. Tichavskỳ and A. Yeredor. Fast approximate joint diagonalization incorporating weight matrices. *Signal Processing, IEEE Transactions on*, 57(3):878–891, 2009.

[10] F. Bouchard, B. Afsari, J. Malick, and M. Congedo. Approximate joint diagonalization with Riemannian optimization on the general linear group. *accepted in SIAM Journal on Matrix Analysis and Applications*.

[11] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, February 1997.

[12] D. Lahat, J.-F. Cardoso, and H. Messer. Second-order multidimensional ica: Performance analysis. *IEEE Transactions on Signal Processing*, 60(9):4598–4610, September 2012.

[13] D. Lahat, J.-F. Cardoso, and H. Messer. Blind separation of multi-dimensional components via subspace decomposition: Performance analysis. *IEEE Transactions on Signal Processing*, 62(11):2894–2905, November 2014.

[14] A. Yeredor. Blind separation of Gaussian sources with general covariance structures: Bounds and optimal estimation. *IEEE Transactions on Signal Processing*, 58(10):5057–5068, 2010.

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

[16] N. Boumal. On intrinsic Cramér-Rao bounds for Riemannian submanifolds and quotient manifolds. *IEEE transactions on signal processing*, 61(7):1809–1821, 2013.

[17] A. Breloy, G. Ginolhac, A. Renaux, and F. Bouchard. Intrinsic cramér–rao bounds for scatter and shape matrices estimation in ces distributions. *IEEE Signal Processing Letters*, 26(2):262–266, February 2019.

[18] L. T. Skovgaard. A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, pages 211–223, 1984.

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

[20] E. Andruchow, G. Larotonda, L. Recht, and A. Varela. The left invariant metric in the general linear group. *Journal of Geometry and Physics*, 86:241–257, 2014.

[21] M. Moakher. Means and averaging in the group of rotations. *SIAM journal on matrix analysis and applications*, 24(1):1–16, 2002.

[22] J. H. Manton. A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups. In *ICARCV 2004 8th Control, Automation, Robotics and Vision Conference, 2004.*, volume 3, pages 2211–2216. IEEE, 2004.