On intrinsic Cramér-Rao bounds for Riemannian submanifolds and quotient manifolds

Nicolas Boumal

Pôle de Mathématiques Appliquées, ICTEAM Institute, Université catholique de Louvain, B-1348, Belgium

nicolas.boumal@uclouvain.be

Abstract

We study Cramér-Rao bounds (CRB's) for estimation problems on Riemannian manifolds. In (S.T. Smith, *Covariance, subspace, and intrinsic Cramér-Rao bounds*, IEEE TSP, 53(5):1610–1630, 2005), the author gives intrinsic CRB's in the form of matrix inequalities relating the covariance of estimators and the Fisher information of estimation problems. We focus on estimation problems whose parameter space $\hat{\mathcal{P}}$ is a Riemannian submanifold or a Riemannian quotient manifold of a parent space $\mathcal{P}$, that is, estimation problems with either deterministic constraints or ambiguities. The CRB's in the aforementioned reference would be expressed w.r.t. bases of the tangent spaces to $\hat{\mathcal{P}}$. In some cases though, it is more convenient to express covariance and Fisher information w.r.t. bases of the tangent spaces to $\mathcal{P}$. We give CRB's w.r.t. such bases expressed in terms of the geodesic distances on the parameter space. The bounds are valid even for singular Fisher information matrices. In an example, we show how the CRB's for a type of sensor network localization problem differ in the presence or absence of anchors, leading to estimation on either submanifolds or quotient manifolds with very different interpretations.

Keywords: Cramér-Rao bounds, CRB, Riemannian manifolds, submanifolds, quotient manifolds, intrinsic bounds, estimation bounds, singular Fisher information matrix, singular FIM, graph Laplacian, sensor network localization, synchronization.

EDICS: SSP-PERF (Performance analysis and bounds)

I. Introduction

We study Cramér-Rao bounds (CRB's) for estimation problems on Riemannian manifolds. In such problems, one would like to estimate a deterministic but unknown parameter θ belonging to a manifold

$\mathcal{P}$, given a measurement x belonging to a probability space $\mathcal{M}$. The measurement x is a random variable whose probability density function is shaped by θ. It is because x is distributed differently for different θ's that sampling (observing) x reveals information about θ.

Such problems arise naturally in camera networks pose estimation [1], angular synchronization [2], covariance matrix estimation and subspace estimation [3] and many other applications, see references therein. Cramér-Rao bounds—which are asymptotic for large SNR [3]—relate the covariance matrix of unbiased estimators to the Fisher information matrix (FIM) of an estimation problem through matrix inequalities. The classical results deal with estimation on Euclidean spaces [4]. More recently, a number of authors have established similar bounds in the manifold setting, see [3], [5] and references therein.

More formally, let $\mathcal{P}$ be a Riemannian manifold and $\mathcal{M}$ be a probability space, i.e., a measurable space with measure μ such that $\mu(\mathcal{M}) = 1$. We consider an estimation problem on the parameter space $\mathcal{P}$ based on measurements in $\mathcal{M}$, such that the probability density function of the measurement given a parameter $\theta \in \mathcal{P}$ is $f(\cdot;\theta) : \mathcal{M} \to \mathbb{R}$. Let $L : \mathcal{P} \to \mathbb{R}$ be the associated log-likelihood function

$$L(\theta) = \log f(x;\theta). \tag{1}$$

The related Fisher information form at θ, $\mathbf{F} : \mathrm{T}_\theta\mathcal{P} \times \mathrm{T}_\theta\mathcal{P} \to \mathbb{R}$, is defined as (all expectations are taken w.r.t. x):

$$\mathbf{F}[u,v] \triangleq \mathbb{E}\left\{\mathrm{D}L(\theta)[u] \cdot \mathrm{D}L(\theta)[v]\right\}, \tag{2}$$

where $\mathrm{D}L(\theta)[u]$ denotes the directional derivative of L at θ along the tangent vector $u \in \mathrm{T}_\theta\mathcal{P}$. The bilinear form $\mathbf{F}$ is symmetric, positive semidefinite. If it is positive definite—that is, $\mathbf{F}[u,u] > 0$ whenever $u \neq 0$—the Cramér-Rao bounds in [3], which make use of the inverse of the matrix representing $\mathbf{F}$, apply.

In this paper, we focus on estimation problems such that $\mathbf{F}$ is *not* positive definite. This situation typically arises when the measurements are not sufficient to determine the parameter, i.e., ambiguities remain. For example, locating a point $p = (x,y,z)$ in space based solely on information about the bearing $p/\|p\|$ is impossible, since nothing is known about the distance between p and the origin. The Fisher information matrix of such a problem would only be positive *semi*definite.

To resolve these ambiguities, one can proceed in at least two ways. Firstly, one can add constraints on θ, based on additional knowledge about the parameter. By restricting the parameter space to $\hat{\mathcal{P}} \subset \mathcal{P}$, a submanifold of $\mathcal{P}$, one may hope that the resulting estimation problem is well-posed. For example, if one knows the distance between p and the origin is 1, one should perform the estimation on $\hat{\mathcal{P}} = \mathbb{S}^2 = \{(x,y,z) : x^2+y^2+z^2 = 1\}$ rather than on $\mathcal{P} = \mathbb{R}^3$. Alternatively, one can recognize that the parameter space is made of equivalence classes, that is, sets of parameters that are equally valid estimators for they give rise to the same measurement distribution. In this scenario, one ends up with an estimation problem on a quotient manifold $\hat{\mathcal{P}} = \mathcal{P}/\sim$, where $\sim$ is an equivalence relation on $\mathcal{P}$. Continuing with our example, all points p with the same bearing $p/\|p\|$ would give rise to the same measurement distribution, hence are indistinguishable and should be grouped into an equivalence class.

The treatment of submanifolds hereafter may also be useful when the FIM is invertible. In that scenario, one is interested in studying the Cramér-Rao bounds of the original problem, and the effect on those bounds caused by incorporating additional knowledge about θ.

The standard way to address these issues would be to work on $\hat{\mathcal{P}}$ directly, writing down Fisher information and covariance with respect to bases of the tangent spaces to $\hat{\mathcal{P}}$, leading to Cramér-Rao bounds according to [3]. However, we argue that the tangent spaces of $\mathcal{P}$ sometimes make more sense to the user: that is why the problem was defined on $\mathcal{P}$ rather than $\hat{\mathcal{P}}$ to begin with. Furthermore, when $\hat{\mathcal{P}}$ is a quotient manifold, its tangent spaces are rather abstract objects to work with. It is hence desirable to have equivalent Cramér-Rao bounds expressed as matrix inequalities w.r.t. bases of tangent spaces of $\mathcal{P}$ instead. This is what the theorems in this communication achieve. We base our work on [3] and derive its consequences for unbiased estimators in the presence of indeterminacies (ambiguities) or under additional constraints.

The case of constrained Cramér-Rao bounds, that is, estimation on Riemannian submanifolds of $\mathbb{R}^d$, was notably addressed in [6]. In their formulation, the authors describe $\hat{\mathcal{P}}$ through a set of equality constraints and they express the covariance in terms of distances in the embedding Euclidean space $\mathbb{R}^d$. In this paper, we consider Riemannian submanifolds of any Riemannian manifold $\mathcal{P}$. Furthermore, for the simple versions of the CRB's, we only require an orthogonal projector from the tangent spaces of $\mathcal{P}$ to those of $\hat{\mathcal{P}}$. More importantly, our covariance matrix is expressed in terms of the Riemannian, or geodesic, distance on $\hat{\mathcal{P}}$, which may be more natural for a number of applications.

The case of CRB's for estimation problems with singular FIM due to indeterminacies, that is, estimation on Riemannian quotient manifolds, was addressed in [7]. In that paper, the authors give a geometric interpretation for the kernel of the FIM and propose a CRB-type bound they name IVLB [5] for the variance of unbiased estimators for such problems. In their bound, the pseudoinverse of the FIM appears naturally and the possible curvature of $\hat{\mathcal{P}}$ is captured through a single number: an upperbound on the sectional curvatures of $\hat{\mathcal{P}}$. In comparison, since we base our results on [3], we propose bounds for the whole covariance matrix (the trace of which coincides with the variance). The pseudoinverse of the FIM also appears naturally. The additional curvature terms in the CRB (Section IV) take the whole Riemannian curvature tensor into account. This is especially useful when $\hat{\mathcal{P}}$ is flat or almost flat in most directions but has significant curvature in a few directions, which happens naturally for product spaces. In such scenarios, the IVLB tends to be overly optimistic—hence less informative—because it has to assume maximum curvature in all directions. In comparison, our bounds derived from [3] are able to capture this complex curvature structure.

Let $e = \{e_1, \ldots, e_d\}$ be an orthonormal basis of $\mathrm{T}_\theta\mathcal{P}$ w.r.t. the Riemannian metric $\langle\cdot,\cdot\rangle_\theta$. The Fisher information matrix of the estimation problem on $\mathcal{P}$ w.r.t. the basis e is defined by:

$$(F_e)_{ij} = \mathbf{F}[e_i, e_j] = \mathbb{E}\left\{\mathrm{D}L(\theta)[e_i] \cdot \mathrm{D}L(\theta)[e_j]\right\}. \tag{3}$$

The covariance matrix C_e w.r.t. the basis e will be defined separately for the submanifold (Section II) and the quotient manifold (Section III) cases, then F_e and C_e will be linked through matrix inequalities. At first, we will neglect curvature terms due to the possible curvature of $\hat{\mathcal{P}}$, then we will establish the CRB's including curvature terms (Section IV). Finally, we will illustrate the usage of these theorems through an example (Section V).

II. Riemannian submanifolds

Let us consider a related estimation problem on the space $\hat{\mathcal{P}} \subset \mathcal{P}$, a Riemannian submanifold of $\mathcal{P}$, such that $\theta \in \hat{\mathcal{P}}$ and for which the log-likelihood function $\hat{L} = L|_{\hat{\mathcal{P}}}$ is the restriction of L to $\hat{\mathcal{P}}$. This situation arises when one adds supplementary constraints on the parameter θ. For example, some of the target parameters are known, or related. The Fisher information is simply $\hat{\mathbf{F}} = \mathbf{F}|_{\mathrm{T}_\theta\hat{\mathcal{P}}\times\mathrm{T}_\theta\hat{\mathcal{P}}}$. We assume $\hat{\mathbf{F}}$ is invertible, i.e., the added constraints fix possible ambiguities in the estimation problem. Figure 1 depicts the situation.

Let $\hat{\theta} : \mathcal{M} \to \hat{\mathcal{P}}$ be any unbiased estimator for our problem. We define the covariance matrix of $\hat{\theta}$ w.r.t. the basis e as:

$$(C_e)_{ij} = \mathbb{E}\left\{\left\langle \mathrm{Log}_\theta(\hat{\theta}), e_i\right\rangle_\theta \cdot \left\langle \mathrm{Log}_\theta(\hat{\theta}), e_j\right\rangle_\theta\right\}, \tag{4}$$

where $\mathrm{Log}_\theta : \hat{\mathcal{P}} \to \mathrm{T}_\theta\hat{\mathcal{P}}$ is the logarithmic map at θ on $\hat{\mathcal{P}}$ [8]. For example, on a Euclidean space, $\mathrm{Log}_\theta(\hat{\theta}) = \hat{\theta} - \theta$. Our goal is to link C_e and F_e through a matrix inequality.

Let $\hat{e} = \{\hat{e}_1, \ldots, \hat{e}_{\hat{d}}\}$ be an orthonormal basis of $\mathrm{T}_\theta\hat{\mathcal{P}} \subset \mathrm{T}_\theta\mathcal{P}$ w.r.t. the Riemannian metric $\langle\cdot,\cdot\rangle_\theta$. Let E be the $\hat{d} \times d$ matrix such that $E_{ij} = \langle \hat{e}_i, e_j\rangle_\theta$. E is orthonormal: $EE^\top = I_{\hat{d}}$, but in general, $P_e \triangleq E^\top E \neq I_d$. Furthermore, let $P_\theta : \mathrm{T}_\theta\mathcal{P} \to \mathrm{T}_\theta\hat{\mathcal{P}}$ be the orthonormal projector onto $\mathrm{T}_\theta\hat{\mathcal{P}}$. Obviously, P_e is the matrix representation of P_θ w.r.t. the basis e, that is: $\langle P_\theta e_i, e_j\rangle = (P_e)_{ij}$.

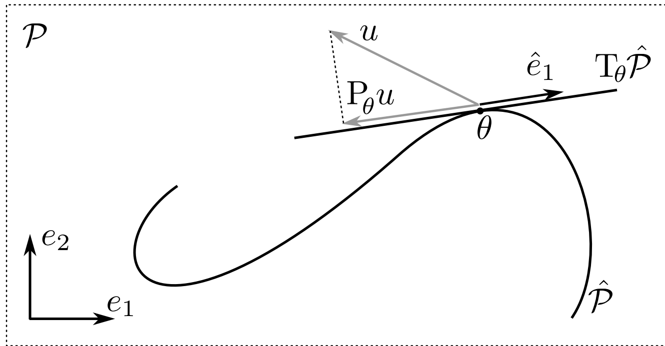

Fig. 1. $\hat{\mathcal{P}}$ is a Riemannian submanifold of $\mathcal{P}$. We consider estimation problems for which the parameter to estimate is θ, a point of $\hat{\mathcal{P}}$. In this drawing, for simplicity, we chose $\mathcal{P} = \mathbb{R}^2$. The vectors $e = (e_1, e_2)$ form an orthonormal basis of $\mathrm{T}_\theta\mathcal{P} \equiv \mathcal{P}$, while $\hat{e} = (\hat{e}_1)$ is an orthonormal basis of the tangent space $\mathrm{T}_\theta\hat{\mathcal{P}}$. The operator P_θ projects vectors of $\mathrm{T}_\theta\mathcal{P}$ orthogonally to $\mathrm{T}_\theta\hat{\mathcal{P}}$. We express the Cramér-Rao bounds for such problems in terms of the basis e, which at times may be more convenient than defining a basis $\hat{e}$ for each point θ and for each different definition of $\hat{\mathcal{P}}$.

A direct application of the CRB's in [3] to the estimation problem on $\hat{\mathcal{P}}$ would link the covariance matrix $C_{\hat{e}}$ of $\hat{\theta}$ and the inverse Fisher information matrix $\hat{F}_{\hat{e}}^{-1}$ w.r.t. the basis $\hat{e}$. More precisely,

$$\begin{aligned}
(C_{\hat{e}})_{ij} &= \mathbb{E}\left\{\left\langle \mathrm{Log}_\theta(\hat{\theta}), \hat{e}_i\right\rangle_\theta \cdot \left\langle \mathrm{Log}_\theta(\hat{\theta}), \hat{e}_j\right\rangle_\theta\right\}, \\
(\hat{F}_{\hat{e}})_{ij} &= \hat{\mathbf{F}}[\hat{e}_i, \hat{e}_j] = \mathbf{F}[\hat{e}_i, \hat{e}_j], \\
C_{\hat{e}} &\succeq \hat{F}_{\hat{e}}^{-1} \quad \text{(neglecting curvature terms).}
\end{aligned} \tag{5}$$

We argue that it is sometimes convenient to work with C_e and F_e directly, to avoid the necessity to define and work with the basis $\hat{e}$. This is what the next theorem achieves, right after we establish a technical lemma.

Lemma 1. *Let $E \in \mathbb{R}^{\hat{d}\times d}$, $A \in \mathbb{R}^{d\times d}$, $B \in \mathbb{R}^{\hat{d}\times\hat{d}}$, with $\hat{d} \leq d$, $A = A^\top$, $B = B^\top$ and $EE^\top = I_{\hat{d}}$, i.e., E is orthonormal. Further assume that* $\ker E \subset \ker A$. *Then,*

$$EAE^\top \succeq B \quad \Rightarrow \quad A \succeq E^\top BE. \tag{6}$$

Proof: Since $\mathbb{R}^d = \mathrm{im}\, E^\top \oplus \ker E$, for all $x \in \mathbb{R}^d$, there exist unique vectors $y \in \mathbb{R}^{\hat{d}}$ and $z \in \mathbb{R}^d$

such that $x = E^\top y + z$ and $Ez = 0$. It follows that:

$$\begin{aligned} x^\top A x &= y^\top E A E^\top y + z^\top A z + 2 y^\top E A z \\ (\text{since } Ez = 0 \Rightarrow Az = 0) \quad &= y^\top E A E^\top y \\ (\text{since } EAE^\top \succeq B) \quad &\geq y^\top B y \\ (\text{since } Ex = EE^\top y + Ez = y) \quad &= x^\top E^\top B E x. \end{aligned}$$

This holds for all x, hence $A \succeq E^\top B E$. ∎

Theorem 2 (CRB on submanifolds)**.** *Given any unbiased estimator $\hat{\theta}$ for the estimation problem on the Riemannian submanifold $\hat{\mathcal{P}}$, the $d \times d$ covariance matrix C_e* (4) *and the $d \times d$ Fisher information matrix F_e* (3) *obey the matrix inequality (neglecting curvature terms of $\hat{\mathcal{P}}$ and assuming $\operatorname{rank}(P_e F_e P_e) = \hat{d}$):*

$$C_e \succeq (P_e F_e P_e)^\dagger, \tag{7}$$

where the $d \times d$ matrix $P_e = E^\top E$ is the orthogonal projector from $\mathrm{T}_\theta \mathcal{P}$ to $\mathrm{T}_\theta \hat{\mathcal{P}}$ w.r.t. the basis e and $\dagger$ denotes Moore-Penrose inversion. Furthermore, the spectrum of $(P_e F_e P_e)^\dagger$ is the spectrum of $\hat{F}_{\hat{e}}^{-1}$ with $d - \hat{d}$ additional zeroes. In particular:

$$\operatorname{trace}(C_e) = \operatorname{trace}(C_{\hat{e}}) \geq \operatorname{trace}(\hat{F}_{\hat{e}}^{-1}) = \operatorname{trace}((P_e F_e P_e)^\dagger).$$

Proof: Since $\hat{\theta} \in \hat{\mathcal{P}}$, $\operatorname{Log}_\theta(\hat{\theta}) \in \mathrm{T}_\theta \hat{\mathcal{P}}$. Consequently, for all $u \in \mathrm{T}_\theta \mathcal{P}$, $\left\langle \operatorname{Log}_\theta(\hat{\theta}), u \right\rangle_\theta = \left\langle \operatorname{Log}_\theta(\hat{\theta}), P_\theta u \right\rangle_\theta$, where $P_\theta u$ is the orthogonal projection of u on $\mathrm{T}_\theta \hat{\mathcal{P}}$. The orthogonal projection of the basis vector e_i on $\mathrm{T}_\theta \hat{\mathcal{P}}$ expands in the basis $\hat{e}$ as $P_\theta e_i = \sum_j \langle \hat{e}_j, e_i \rangle_\theta \hat{e}_j = \sum_j E_{ji} \hat{e}_j$. Then, by bilinearity, $(C_e)_{ij} = \sum_{k,\ell} E_{ki} E_{\ell j} (C_{\hat{e}})_{k\ell}$. In matrix form,

$$C_e = E^\top C_{\hat{e}} E. \tag{8}$$

Since $EE^\top = I_{\hat{d}}$, it also holds that $C_{\hat{e}} = E C_e E^\top$. The vectors of $\hat{e}$ expand in the basis e as $\hat{e}_i = \sum_j \langle \hat{e}_i, e_j \rangle_\theta e_j = \sum_j E_{ij} e_j$. By bilinearity again, $(\hat{F}_{\hat{e}})_{ij} = \sum_{k,\ell} E_{ik} E_{j\ell} (F_e)_{k\ell}$. In matrix form,

$$\hat{F}_{\hat{e}} = E F_e E^\top. \tag{9}$$

Notice that the assumption $\operatorname{rank}(P_e F_e P_e) = \hat{d}$ is equivalent to the assumption that $\hat{F}_{\hat{e}}$ is invertible. Then, substituting in (5), we find $E C_e E^\top \succeq (E F_e E^\top)^{-1}$. Since $\ker C_e = \ker(E^\top C_{\hat{e}} E) \supset \ker E$, Lemma 1 applies and it follows that:

$$C_e \succeq E^\top (E F_e E^\top)^{-1} E. \tag{10}$$

Finally, from the definition of Moore-Penrose pseudoinverse, it is easily checked that

$$E^\top(EF_eE^\top)^{-1}E = (E^\top EF_eE^\top E)^\dagger. \tag{11}$$

Since $P_e = E^\top E$, this concludes the proof of the main part.

We now establish the spectrum property. Since $\hat{F}_{\hat{e}}^{-1}$ is symmetric positive definite, there exist a diagonal matrix D and an orthogonal matrix U of size $\hat{d}\times\hat{d}$ such that $\hat{F}_{\hat{e}}^{-1} = UDU^\top$. Hence,

$$(P_eF_eP_e)^\dagger = E^\top UDU^\top E = V\begin{pmatrix} D & \\ & 0\end{pmatrix}V^\top, \tag{12}$$

with $V = \begin{pmatrix} E^\top U & (E^\top U)^\perp\end{pmatrix}$ a $d\times d$ orthogonal matrix. The trace property follows easily:

$$\operatorname{trace}(C_e) = \operatorname{trace}(E^\top C_{\hat{e}}E) = \operatorname{trace}(C_{\hat{e}}) \geq \operatorname{trace}(\hat{F}_{\hat{e}}^{-1}) = \operatorname{trace}((P_eF_eP_e)^\dagger). \tag{13}$$

■

The trace property is especially interesting, as it bounds the variance of the estimator $\hat{\theta}$, expressed w.r.t. the Riemannian distance dist on $\hat{\mathcal{P}}$:

$$\operatorname{trace}(C_e) = \operatorname{trace}(C_{\hat{e}}) = \mathbb{E}\left\{\|\operatorname{Log}_\theta(\hat{\theta})\|^2\right\} = \mathbb{E}\left\{\operatorname{dist}^2(\theta,\hat{\theta})\right\} \triangleq \operatorname{var}_{\hat{\theta}}(\theta). \tag{14}$$

Here is one way of interpreting the bound (7). Expand the random error vector $\operatorname{Log}_\theta(\hat{\theta}) = \sum_i x_ie_i$ with random coefficients x_i. From the definition, $(C_e)_{ii} = \mathbb{E}\left\{x_i^2\right\}$. Then, equation (7) implies $\mathbb{E}\left\{x_i^2\right\} \geq (P_eF_eP_e)_{ii}^\dagger$, which is a lower-bound on how well the coefficient x_i can be estimated.

Notice that it is not necessary to explicitly construct a basis $\hat{e}$ in order to use Theorem 2. Indeed, the orthogonal projector P_e is often easy to compute without requiring an explicit factorization as $E^\top E$. For example, the orthogonal projector from $\mathbb{R}^3$ onto the tangent space to the sphere $\mathbb{S}^2$ at θ, denoted $\mathrm{T}_\theta\mathbb{S}^2$, w.r.t. the canonical basis of $\mathbb{R}^3$ is simply $P_e = I_3 - \theta\theta^\top$, where I_3 is the 3×3 identity matrix. Conversely, because of the hairy ball theorem, it is impossible to define bases $\hat{e}$ of $\mathrm{T}_\theta\mathbb{S}^2$ for all θ in a smooth way, making it rather inconvenient to work with such bases.

III. Riemannian quotient manifolds

Whenever two parameters $\theta, \theta' \in \mathcal{P}$ give rise to the same measurement distribution, they are indistinguishable, in the sense that no argument based on the observed measurement can be used to favor one parameter over the other as estimator. This observation motivates the definition of the following equivalence relation:

$$\theta \sim \theta' \Leftrightarrow f(\cdot,\theta) \equiv f(\cdot,\theta') \text{ a.e. on } \mathcal{M}. \tag{15}$$

The quotient space $\hat{\mathcal{P}} = \mathcal{P}/\sim$—that is, the set of equivalence classes—then becomes the natural parameter space on which the estimation should be performed. Figures 2 and 3, courtesy of the authors of [7], depict the concept of quotient manifold and of the related basic objects we introduce hereafter, namely submersions and horizontal/vertical spaces.

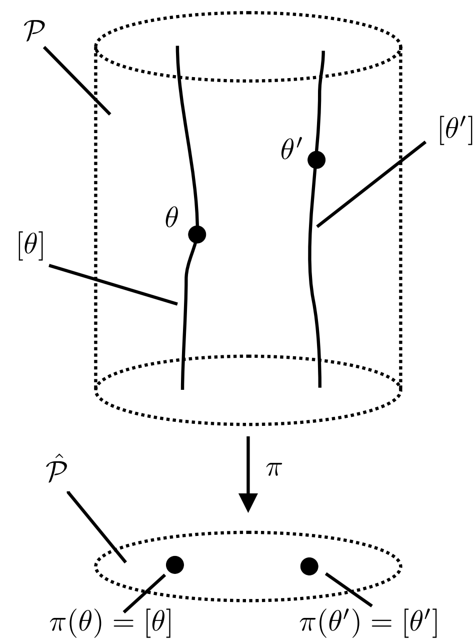

Fig. 2. The parameter space $\mathcal{P}$ is partitioned into equivalence classes, called fibers. The Riemannian submersion π maps each $\theta \in \mathcal{P}$ to its corresponding equivalence class $[\theta] \in \hat{\mathcal{P}}$. The space of equivalence classes is the quotient space $\hat{\mathcal{P}} = \mathcal{P}/\sim$, also a Riemannian manifold. *Figure courtesy of [7].*

We now consider the mapping from $\mathcal{P}$ to $\hat{\mathcal{P}}$,

$$\pi : \mathcal{P} \to \hat{\mathcal{P}} : \theta \mapsto \pi(\theta) = [\theta] \triangleq \{\theta' \in \mathcal{P} : \theta' \sim \theta\}, \tag{16}$$

and concentrate on the case where π is a Riemannian submersion [8][9]. That is, $\hat{\mathcal{P}}$ is a Riemannian quotient manifold of $\mathcal{P}$. In particular, $[\theta]$ is a Riemannian submanifold of $\mathcal{P}$ (a *fiber*). The log-likelihood function $\hat{L} : \hat{\mathcal{P}} \to \mathbb{R}$ is well-defined by $\hat{L}([\theta]) \triangleq L(\theta)$.

The tangent space to $[\theta]$ at θ, named the vertical space V_θ, is a subspace of the tangent space $\mathrm{T}_\theta\mathcal{P}$. The orthogonal complement of the vertical space, named the horizontal space H_θ, is such that $\mathrm{T}_\theta\mathcal{P} = \mathrm{H}_\theta \oplus \mathrm{V}_\theta$. The pushforward $\mathrm{D}\pi(\theta) : \mathrm{T}_\theta\mathcal{P} \to \mathrm{T}_{[\theta]}\hat{\mathcal{P}}$ of a Riemannian submersion induces a metric on the abstract tangent space $\mathrm{T}_{[\theta]}\hat{\mathcal{P}}$:

$$\forall u, v \in \mathrm{H}_\theta, \quad \langle \mathrm{D}\pi(\theta)[u], \mathrm{D}\pi(\theta)[v] \rangle_{[\theta]} \triangleq \langle u, v \rangle_\theta \,. \tag{17}$$

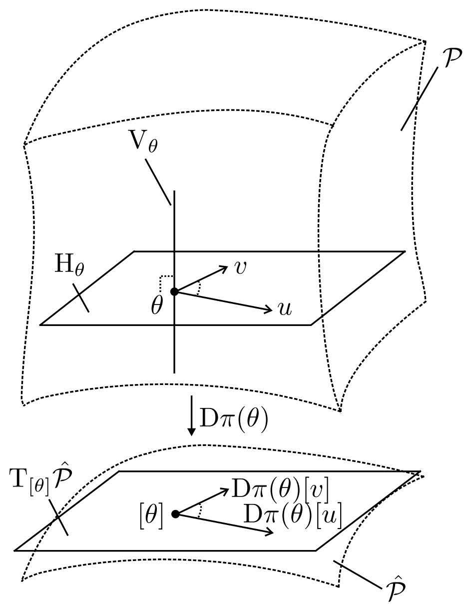

Fig. 3. The tangent space at θ in $\mathcal{P}$, $\mathrm{T}_\theta\mathcal{P}$, is the direct sum of a horizontal space H_θ and a vertical space V_θ. The differential of π, noted $\mathrm{D}\pi(\theta)$, is an isometry between H_θ and the abstract tangent space $\mathrm{T}_{[\theta]}\hat{\mathcal{P}}$. This makes it convenient to represent abstract tangent vectors to $\hat{\mathcal{P}}$ as horizontal vectors. *Figure courtesy of [7].*

The definition of Riemannian submersion ensures that this is well-defined [9]. We mention two useful properties:

$$\ker \mathrm{D}\pi(\theta) = \mathrm{V}_\theta, \quad \text{and} \tag{18}$$

$$\mathrm{D}\pi(\theta)|_{\mathrm{H}_\theta} : \mathrm{H}_\theta \to \mathrm{T}_{[\theta]}\hat{\mathcal{P}} \text{ is an isometry.} \tag{19}$$

Let $[\hat{\theta}] : \mathcal{M} \to \hat{\mathcal{P}}$ be any unbiased estimator for our problem. We define the covariance matrix of $[\hat{\theta}]$ w.r.t. the basis e as:

$$(C_e)_{ij} = \mathbb{E}\left\{\langle \xi, e_i\rangle_\theta \cdot \langle \xi, e_j\rangle_\theta\right\}, \text{ with}$$
$$\xi = (\mathrm{D}\pi(\theta)|_{\mathrm{H}_\theta})^{-1}[\mathrm{Log}_{[\theta]}([\hat{\theta}])]. \tag{20}$$

The error vector ξ is the shortest horizontal vector at θ such that $\mathrm{Exp}_\theta(\xi) \in [\hat{\theta}]$.

Let $\hat{e} = (\hat{e}_1, \ldots, \hat{e}_{\hat{d}})$ be an orthonormal basis of $\mathrm{T}_{[\theta]}\hat{\mathcal{P}}$. A direct application of the CRB's in [3] to the estimation problem on $\hat{\mathcal{P}}$ would link the covariance matrix $C_{\hat{e}}$ of $[\hat{\theta}]$ and the inverse Fisher information matrix $\hat{F}_{\hat{e}}^{-1}$ w.r.t. the basis $\hat{e}$. More precisely,

$$
\begin{aligned}
(C_{\hat{e}})_{ij} &= \mathbb{E}\left\{ \left\langle \mathrm{Log}_{[\theta]}([\hat{\theta}]), \hat{e}_i \right\rangle_{[\theta]} \cdot \left\langle \mathrm{Log}_{[\theta]}([\hat{\theta}]), \hat{e}_j \right\rangle_{[\theta]} \right\}, \\
(\hat{F}_{\hat{e}})_{ij} &= \hat{\mathbf{F}}[\hat{e}_i, \hat{e}_j] = \mathbb{E}\left\{ \mathrm{D}\hat{L}([\theta])[\hat{e}_i] \cdot \mathrm{D}\hat{L}([\theta])[\hat{e}_j] \right\}, \\
C_{\hat{e}} &\succeq \hat{F}_{\hat{e}}^{-1} \quad \text{(neglecting curvature terms).}
\end{aligned} \tag{21}
$$

Since $\mathrm{T}_{[\theta]}\hat{\mathcal{P}}$ is an abstract space, we argue that it is often convenient to work with the more concrete objects C_e and F_e instead.

Theorem 3 (CRB on quotient manifolds)**.** *Given any unbiased estimator $[\hat{\theta}]$ for the estimation problem on the Riemannian quotient manifold $\hat{\mathcal{P}} = \mathcal{P}/\sim$, the $d \times d$ covariance matrix C_e* (20) *and the $d \times d$ Fisher information matrix F_e* (3) *obey the matrix inequality (neglecting curvature terms of $\hat{\mathcal{P}}$ and assuming $\mathrm{rank}(F_e) = \hat{d}$):*

$$
C_e \succeq F_e^{\dagger}, \tag{22}
$$

where $\dagger$ denotes Moore-Penrose inversion. Furthermore, the spectrum of $F_e^{\dagger}$ is the spectrum of $\hat{F}_{\hat{e}}^{-1}$ with $d - \hat{d}$ additional zeroes. In particular:

$$
\mathrm{trace}(C_e) = \mathrm{trace}(C_{\hat{e}}) \geq \mathrm{trace}(\hat{F}_{\hat{e}}^{-1}) = \mathrm{trace}(F_e^{\dagger}). \tag{23}
$$

Proof: It is convenient to introduce the orthonormal basis of H_θ related to $\hat{e}$ as $\tilde{e} = (\tilde{e}_1, \ldots, \tilde{e}_{\hat{d}})$, with $\hat{e}_i = \mathrm{D}\pi(\theta)[\tilde{e}_i]$. The $\hat{d} \times d$ matrix E such that $E_{ij} = \langle \tilde{e}_i, e_j \rangle_\theta$ will prove useful. E is orthonormal: $EE^\top = I_{\hat{d}}$, but in general, $E^\top E \neq I_d$.

Let us denote the orthogonal projection of $u \in \mathrm{T}_\theta\mathcal{P}$ onto the horizontal space H_θ as $P\mathrm{H}_\theta u$. Since $\xi = (\mathrm{D}\pi(\theta)|_{\mathrm{H}_\theta})^{-1}[\mathrm{Log}_{[\theta]}([\hat{\theta}])]$ is a horizontal vector, $\langle \xi, u \rangle_\theta = \langle \xi, P\mathrm{H}_\theta u \rangle$. Furthermore, $\mathrm{D}\pi(\theta)[P\mathrm{H}_\theta u] = \mathrm{D}\pi(\theta)[u]$. Then, using the fact that $\mathrm{D}\pi(\theta)|_{\mathrm{H}_\theta}$ is an isometry, it follows that

$$
(C_e)_{ij} = \mathbb{E}\left\{ \langle \xi, e_i \rangle_\theta \cdot \langle \xi, e_j \rangle_\theta \right\} \tag{24}
$$

$$
= \mathbb{E}\left\{ \langle \xi, P\mathrm{H}_\theta e_i \rangle_\theta \cdot \langle \xi, P\mathrm{H}_\theta e_j \rangle_\theta \right\} \tag{25}
$$

$$
= \mathbb{E}\left\{ \left\langle \mathrm{Log}_{[\theta]}([\hat{\theta}]), \mathrm{D}\pi(\theta)[e_i] \right\rangle_{[\theta]} \cdot \left\langle \mathrm{Log}_{[\theta]}([\hat{\theta}]), \mathrm{D}\pi(\theta)[e_j] \right\rangle_{[\theta]} \right\}. \tag{26}
$$

The vector $\mathrm{D}\pi(\theta)[e_i] \in \mathrm{T}_{[\theta]}\hat{\mathcal{P}}$ expands in the basis $\hat{e}$ as $\mathrm{D}\pi(\theta)[e_i] = \sum_j E_{ji}\hat{e}_j$. Indeed,

$$
\langle \mathrm{D}\pi(\theta)[e_i], \hat{e}_j \rangle_{[\theta]} = \langle \mathrm{D}\pi(\theta)[e_i], \mathrm{D}\pi(\theta)[\tilde{e}_j] \rangle_{[\theta]} = \langle e_i, \tilde{e}_j \rangle_\theta .
$$

It follows that $(C_e)_{ij} = \sum_{k,\ell} E_{ki} E_{\ell j} \, (C_{\hat{e}})_{k\ell}$. In matrix form:

$$C_e = E^\top C_{\hat{e}} E. \tag{27}$$

Since $EE^\top = I_{\hat{d}}$, it also holds that $C_{\hat{e}} = E C_e E^\top$.

We now similarly link F_e and $\hat{F}_{\hat{e}}$. In doing so, we exploit the fact that $\operatorname{grad} L(\theta)$ is a horizontal vector. This stems from the fact that the log-likelihood function L is constant over fibers (equivalence classes).

$$\begin{aligned}
(F_e)_{ij} &= \mathbb{E}\left\{\mathrm{D}L(\theta)[e_i] \cdot \mathrm{D}L(\theta)[e_j]\right\} \\
&= \mathbb{E}\left\{\langle \operatorname{grad} L(\theta), e_i\rangle_\theta \cdot \langle \operatorname{grad} L(\theta), e_j\rangle_\theta\right\} \\
&= \mathbb{E}\left\{\langle \operatorname{grad} L(\theta), P\mathrm{H}_\theta e_i\rangle_\theta \cdot \langle \operatorname{grad} L(\theta), P\mathrm{H}_\theta e_j\rangle_\theta\right\} \\
&\text{(expand } P\mathrm{H}_\theta e_i \text{ and } P\mathrm{H}_\theta e_j \text{ in the basis } \tilde{e}) \\
&= \sum_{k,\ell} E_{ki} E_{\ell j} \, \mathbb{E}\left\{\langle \operatorname{grad} L(\theta), \tilde{e}_k\rangle_\theta \cdot \langle \operatorname{grad} L(\theta), \tilde{e}_\ell\rangle_\theta\right\} \\
&= \sum_{k,\ell} E_{ki} E_{\ell j} \, \mathbb{E}\left\{\langle \mathrm{D}\pi(\theta)[\operatorname{grad} L(\theta)], \hat{e}_k\rangle_{[\theta]} \cdot \langle \mathrm{D}\pi(\theta)[\operatorname{grad} L(\theta)], \hat{e}_\ell\rangle_{[\theta]}\right\} \\
&= \sum_{k,\ell} E_{ki} E_{\ell j} \, \mathbb{E}\left\{\left\langle \operatorname{grad} \hat{L}([\theta]), \hat{e}_k\right\rangle_{[\theta]} \cdot \left\langle \operatorname{grad} \hat{L}([\theta]), \hat{e}_\ell\right\rangle_{[\theta]}\right\} \\
&= \sum_{k,\ell} E_{ki} E_{\ell j} \, \mathbb{E}\left\{\mathrm{D}\hat{L}([\theta])[\hat{e}_k] \cdot \mathrm{D}\hat{L}([\theta])[\hat{e}_\ell]\right\} \\
&= \sum_{k,\ell} E_{ki} E_{\ell j} \, (\hat{F}_{\hat{e}})_{k\ell}.
\end{aligned}$$

In matrix form,

$$F_e = E^\top \hat{F}_{\hat{e}} E. \tag{28}$$

This highlights the fact that $\ker F_e = \ker E$ (since $\hat{F}_{\hat{e}}$ is invertible), which makes sense since $\ker E \simeq \mathrm{V}_\theta$. Again, by orthonormality of E, it also holds that $\hat{F}_{\hat{e}} = E F_e E^\top$. Combining these rules, it follows that:

$$F_e = E^\top E F_e E^\top E. \tag{29}$$

Notice that the assumption $\operatorname{rank}(F_e) = \hat{d}$ is equivalent to the assumption that $\hat{F}_{\hat{e}}$ is invertible. Applying Lemma 1 to the inequality $C_{\hat{e}} \succeq \hat{F}_{\hat{e}}^{-1}$ (21) and using arguments similar to the proof of Theorem 2 finally yields:

$$C_e \succeq E^\top (E F_e E^\top)^{-1} E = (E^\top E F_e E^\top E)^\dagger = F_e^\dagger. \tag{30}$$

The spectrum and trace properties follow directly, see proof of Theorem 2. ■

Again, there is no need to construct bases $\tilde{e}$ or $\hat{e}$ in order to use Theorem 3. Notice that it still holds that $\operatorname{trace}(C_e) = \operatorname{trace}(C_{\hat{e}}) = \mathbb{E}\left\{\|\xi\|_\theta^2\right\} = \mathbb{E}\left\{\operatorname{dist}^2([\theta],[\hat{\theta}])\right\}$, where dist is the Riemannian distance on $\hat{\mathcal{P}}$, since $\mathrm{D}\pi(\theta)|_{\mathrm{H}_\theta}$ is an isometry.

IV. Including curvature terms

The intrinsic Cramér-Rao bounds developed in [3] include special terms accounting for the possible curvature of the parameter space $\hat{\mathcal{P}}$—for a reference on curvature, see [10]. These terms vanish if $\hat{\mathcal{P}}$ is flat, that is, if its Riemannian curvature tensor is zero. When $\hat{\mathcal{P}}$ is not flat, the curvature terms may nevertheless often be neglected for high enough signal to noise ratio. The argument developed in [3] to that end concludes that neglecting the curvature terms is legitimate as soon as estimation errors obey

$$\operatorname{dist}(\theta,\hat{\theta}) \ll \frac{1}{\sqrt{K_{\max}}}, \tag{31}$$

where $K_{\max}$ is an upperbound on the sectional curvatures of $\hat{\mathcal{P}}$ at θ.

Condition (31) involves an upperbound on the sectional curvature of $\hat{\mathcal{P}}$. As a consequence, it may be overly restrictive for parameter spaces which have small curvature in most directions, and large curvature in a few. An important class of such spaces consists of all product manifolds.

As an example, let us consider the problem of estimating $(\theta_1,\ldots,\theta_N) \in \hat{\mathcal{P}} = \mathbb{S}^2 \times\cdots\times \mathbb{S}^2$, the product of N spheres. $\hat{\mathcal{P}}$ has unit curvature along 2-planes pertaining to a single sphere, but zero curvature along all 2-planes spanning exactly two distinct spheres. Of course, $K_{\max} = 1$. If estimating θ_i and θ_j, $i \neq j$, are two independent but identical tasks, one should expect the distribution of $\operatorname{dist}(\theta_i,\hat{\theta}_i)$ to be independent of i. Consequently, $\operatorname{dist}(\theta,\hat{\theta})$ grows as $\sqrt{N}$, whereas $K_{\max}$ remains constant. Hence, condition (31) becomes increasingly restrictive with growing N. Of course, since the N tasks are independent and can be considered separately, the negligibility of the curvature terms should not depend on N, which brings the conclusion that simply describing the curvature of $\hat{\mathcal{P}}$ through $K_{\max}$ may not be enough.

For such parameter spaces, it is necessary to explicitly compute the curvature terms in the intrinsic Cramér-Rao bounds, if only to show that they are indeed negligible at reasonable SNR. We now set out to give versions of theorems 2 and 3 including curvature terms, computable without constructing other bases than e, the basis of $\mathrm{T}_\theta\mathcal{P}$. This will require the Riemannian curvature tensor of $\hat{\mathcal{P}}$. Useful references to look up/compute this tensor are [8, Lemma 3.39, Cor. 3.58, Thm 7.47, Cor. 11.10][10][11].

A. Curvature terms for submanifolds

The random error vector $X \triangleq \operatorname{Log}_\theta(\hat{\theta})$ expands in the basis $\hat{e}$ as $X = \sum_i \hat{x}_i \hat{e}_i$, with $\hat{x}_1,\ldots,\hat{x}_{\hat{d}}$ random variables. Notice that $(C_{\hat{e}})_{ij} = \mathbb{E}\left\{\langle X,\hat{e}_i\rangle_\theta \langle X,\hat{e}_j\rangle_\theta\right\} = \mathbb{E}\left\{\hat{x}_i\hat{x}_j\right\}$. Let $\hat{R}$ be the Riemannian curvature tensor of $\hat{\mathcal{P}}$. The mapping $(u, v, w, z) \in (\mathrm{T}_\theta\hat{\mathcal{P}})^4 \mapsto \langle \hat{R}(u, v)w, z\rangle_\theta$ is linear in its four arguments [10]. Smith introduces the symmetric 2-form $\hat{\mathbf{R}}_{\mathbf{m}} : \mathrm{T}_\theta\hat{\mathcal{P}} \times \mathrm{T}_\theta\hat{\mathcal{P}} \to \mathbb{R}$ defined by [3, eq. (34)]:

$$\hat{\mathbf{R}}_{\mathbf{m}}[\hat{e}_i, \hat{e}_j] = \mathbb{E}\left\{\left\langle \hat{R}(X, \hat{e}_i)\hat{e}_j, X\right\rangle_\theta\right\} \tag{32}$$

$$= \mathbb{E}\left\{\sum_{k,\ell}\left\langle \hat{R}(\hat{e}_k, \hat{e}_i)\hat{e}_j, \hat{e}_\ell\right\rangle_\theta \hat{x}_k\hat{x}_\ell\right\} \tag{33}$$

$$= \sum_{k,\ell}\left\langle \hat{R}(\hat{e}_k, \hat{e}_i)\hat{e}_j, \hat{e}_\ell\right\rangle_\theta (C_{\hat{e}})_{k\ell}. \tag{34}$$

From the latter expression, it is apparent that the entries of the matrix associated to $\hat{\mathbf{R}}_{\mathbf{m}}$ are linear combinations of the entries of $C_{\hat{e}}$. Generalizing this to any matrix, the following linear map is defined:

$$\hat{R}_m : \mathbb{R}^{\hat{d}\times\hat{d}} \to \mathbb{R}^{\hat{d}\times\hat{d}} : M \mapsto \hat{R}_m(M), \text{ with}$$

$$(\hat{R}_m(M))_{ij} = \sum_{k,\ell}\left\langle \hat{R}(\hat{e}_k, \hat{e}_i)\hat{e}_j, \hat{e}_\ell\right\rangle_\theta M_{k\ell}. \tag{35}$$

Under appropriate assumptions, the CRB with curvature terms in [3] then reads

$$C_{\hat{e}} \succeq \hat{F}_{\hat{e}}^{-1} - \frac{1}{3}\left(\hat{R}_m(\hat{F}_{\hat{e}}^{-1})\hat{F}_{\hat{e}}^{-1} + \hat{F}_{\hat{e}}^{-1}\hat{R}_m(\hat{F}_{\hat{e}}^{-1})\right). \tag{36}$$

In order to provide an equivalent of (36) only referencing the basis e, we introduce the following symmetric 2-form:

$$\mathbf{R}_{\mathbf{m}}[e_i, e_j] \triangleq \hat{\mathbf{R}}_{\mathbf{m}}[P_\theta e_i, P_\theta e_j]. \tag{37}$$

Notice that, since $X \in \mathrm{T}_\theta\hat{\mathcal{P}}$, we have $X = P_\theta X$. Expanding in the basis e, $X = \sum_i x_i e_i = \sum_i x_i P_\theta e_i$ with random variables $x_1, \ldots, x_d$ and $(C_e)_{ij} = \mathbb{E}\{x_i x_j\}$. It follows that:

$$\mathbf{R}_{\mathbf{m}}[e_i, e_j] = \mathbb{E}\left\{\left\langle \hat{R}(X, P_\theta e_i)P_\theta e_j, X\right\rangle_\theta\right\} \tag{38}$$

$$= \sum_{k,\ell}\left\langle \hat{R}(P_\theta e_k, P_\theta e_i)P_\theta e_j, P_\theta e_\ell\right\rangle_\theta (C_e)_{k\ell}. \tag{39}$$

From there, we introduce the following linear map:

$$R_m : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d} : M \mapsto R_m(M), \text{ with}$$

$$(R_m(M))_{ij} = \sum_{k,\ell}\left\langle \hat{R}(P_\theta e_k, P_\theta e_i)P_\theta e_j, P_\theta e_\ell\right\rangle_\theta M_{k\ell}. \tag{40}$$

Riemannian curvature is often specified by a formula for $\langle R(u, v)v, u\rangle$. Hence the standard polarization identity for symmetric bilinear forms may be useful to compute $\mathbf{R}_{\mathbf{m}}$:

$$4\mathbf{R}_{\mathbf{m}}[e_i, e_j] = \mathbf{R}_{\mathbf{m}}[e_i + e_j, e_i + e_j] - \mathbf{R}_{\mathbf{m}}[e_i - e_j, e_i - e_j].$$

We use the linear maps R_m and $\hat{R}_m$ in the following theorem:

Theorem 4 (CRB on submanifolds, with curvature)**.** *(Continued from Theorem 2). Including terms due to the possible curvature of $\hat{\mathcal{P}}$, the covariance matrix C_e* (4) *of any unbiased estimator $\hat{\theta} : \mathcal{M} \to \hat{\mathcal{P}}$ and the Fisher information matrix F_e* (3) *w.r.t. the orthonormal basis e of $\mathrm{T}_\theta\mathcal{P}$ obey the following matrix inequality (assuming $\mathrm{rank}(P_e F_e P_e) = \hat{d}$):*

$$C_e \succeq \tilde{F}_e^\dagger - \frac{1}{3}\left(R_m(\tilde{F}_e^\dagger)\tilde{F}_e^\dagger + \tilde{F}_e^\dagger R_m(\tilde{F}_e^\dagger)\right), \tag{41}$$

where $\tilde{F}_e = P_e F_e P_e$ and $R_m : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ is as defined by (40).

Proof: We start from the CRB w.r.t. the basis $\hat{e}$ (36):

$$C_{\hat{e}} \succeq \hat{F}_{\hat{e}}^{-1} - \frac{1}{3}\left(\hat{R}_m(\hat{F}_{\hat{e}}^{-1})\hat{F}_{\hat{e}}^{-1} + \hat{F}_{\hat{e}}^{-1}\hat{R}_m(\hat{F}_{\hat{e}}^{-1})\right). \tag{42}$$

By expanding the projections $P_\theta e_i = \sum_j \langle \hat{e}_j, e_i \rangle \hat{e}_j = \sum_j E_{ji}\hat{e}_j$ and exploiting the linearity of $\left\langle \hat{R}(u,v)w, z\right\rangle_\theta$ in its four arguments, the matrix relation below comes forth:

$$\forall M \in \mathbb{R}^{d\times d}, \quad R_m(M) = E^\top \hat{R}_m(EME^\top)E. \tag{43}$$

From the proof of Theorem 2, recall that $C_{\hat{e}} = EC_eE^\top$ and $\hat{F}_{\hat{e}}^{-1} = E(P_eF_eP_e)^\dagger E^\top$. The relation (43) yields $\hat{R}_m(\hat{F}_{\hat{e}}^{-1}) = ER_m((P_eF_eP_e)^\dagger)E^\top$. Substituting in the CRB gives:

$$EC_eE^\top \succeq E\left(\tilde{F}_e^\dagger - \frac{1}{3}\left(R_m(\tilde{F}_e^\dagger)\tilde{F}_e^\dagger + \tilde{F}_e^\dagger R_m(\tilde{F}_e^\dagger)\right)\right)E^\top,$$

where we used the fact that $R_m(M)P_e = P_eR_m(M) = R_m(M)$, which is easily established from (43). Lemma 1 applies and concludes the proof, since $P_e(P_eF_eP_e)^\dagger P_e = (P_eF_eP_e)^\dagger$. ■

B. Curvature terms for quotient manifolds

We follow the same line of thought as for submanifolds. The random error vector $X \triangleq \mathrm{Log}_{[\theta]}([\hat{\theta}])$ expands in the basis $\hat{e}$ as $X = \sum_i \hat{x}_i\hat{e}_i$, with $\hat{x}_1, \ldots, \hat{x}_{\hat{d}}$ random variables and $(C_{\hat{e}})_{ij} = \mathbb{E}\{\hat{x}_i\hat{x}_j\}$. Let $\hat{R}$ be the Riemannian curvature tensor of $\hat{\mathcal{P}}$. We consider $\hat{\mathbf{R}}_\mathbf{m} : \mathrm{T}_{[\theta]}\hat{\mathcal{P}} \times \mathrm{T}_{[\theta]}\hat{\mathcal{P}} \to \mathbb{R}$ defined by:

$$\begin{aligned}\hat{\mathbf{R}}_\mathbf{m}[\hat{e}_i, \hat{e}_j] &= \mathbb{E}\left\{\left\langle \hat{R}(X,\hat{e}_i)\hat{e}_j, X\right\rangle_{[\theta]}\right\} &(44)\\ &= \sum_{k,\ell}\left\langle \hat{R}(\hat{e}_k,\hat{e}_i)\hat{e}_j, \hat{e}_\ell\right\rangle_{[\theta]}(C_{\hat{e}})_{k\ell}. &(45)\end{aligned}$$

A linear map on $\hat{d} \times \hat{d}$ matrices follows:

$$\hat{R}_m : \mathbb{R}^{\hat{d}\times\hat{d}} \to \mathbb{R}^{\hat{d}\times\hat{d}} : M \mapsto \hat{R}_m(M), \text{ with}$$
$$(\hat{R}_m(M))_{ij} = \sum_{k,\ell} \left\langle \hat{R}(\hat{e}_k, \hat{e}_i)\hat{e}_j, \hat{e}_\ell \right\rangle_{[\theta]} M_{k\ell}. \tag{46}$$

Under appropriate assumptions, the CRB (36) holds. To express it only referencing the basis e, we introduce the following symmetric 2-form:

$$\mathbf{R_m}[e_i, e_j] \triangleq \hat{\mathbf{R}}_\mathbf{m} \left[\mathrm{D}\pi(\theta)[e_i], \mathrm{D}\pi(\theta)[e_j]\right]. \tag{47}$$

Let $\xi = (\mathrm{D}\pi(\theta)|_{\mathrm{H}_\theta})^{-1}[X]$ be the unique horizontal vector such that $\mathrm{D}\pi(\theta)[\xi] = X$. Expanding ξ in the basis e as $\xi = \sum_i x_i e_i$, we find $X = \sum_i x_i \mathrm{D}\pi(\theta)[e_i]$ with random variables $x_1, \ldots, x_d$ and $(C_e)_{ij} = \mathbb{E}\left\{x_i x_j\right\}$. It follows that:

$$\mathbf{R_m}[e_i, e_j] = \sum_{k,\ell} \left\langle \hat{R}\left(\mathrm{D}\pi(\theta)[e_k], \mathrm{D}\pi(\theta)[e_i]\right) \mathrm{D}\pi(\theta)[e_j], \mathrm{D}\pi(\theta)[e_\ell] \right\rangle_{[\theta]} (C_e)_{k\ell}.$$

From there, we introduce the following linear map:

$$R_m : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d} : M \mapsto R_m(M), \text{ with} \tag{48}$$

$$(R_m(M))_{ij} = \sum_{k,\ell} \left\langle \hat{R}(\mathrm{D}\pi(\theta)[e_k], \mathrm{D}\pi(\theta)[e_i])\mathrm{D}\pi(\theta)[e_j], \mathrm{D}\pi(\theta)[e_\ell] \right\rangle_{[\theta]} M_{k\ell}.$$

Theorem 5 (CRB on quotient manifolds, with curvature)**.** *(Continued from Theorem 3). Including terms due to the possible curvature of $\hat{\mathcal{P}}$, the covariance matrix C_e* (20) *of any unbiased estimator $\hat{\theta} : \mathcal{M} \to \hat{\mathcal{P}}$ and the Fisher information matrix F_e* (3) *w.r.t. the orthonormal basis e of $\mathrm{T}_\theta\mathcal{P}$ obey the following matrix inequality (assuming $\mathrm{rank}(F_e) = \hat{d}$):*

$$C_e \succeq F_e^\dagger - \frac{1}{3}\left(R_m(F_e^\dagger)F_e^\dagger + F_e^\dagger R_m(F_e^\dagger)\right), \tag{49}$$

where $R_m : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ is as defined by (48)*.*

Proof: The proof is very similar to that of Theorem 4. We start from the CRB w.r.t. the basis $\hat{e}$ (36). Expanding $\mathrm{D}\pi(\theta)[e_i] = \mathrm{D}\pi(\theta)[P\mathrm{H}_\theta e_i] = \sum_j \langle \tilde{e}_j, e_i\rangle \mathrm{D}\pi(\theta)[\tilde{e}_j] = \sum_j E_{ji}\hat{e}_j$ and exploiting linearity of $\langle \hat{R}(\cdot,\cdot)\cdot,\cdot\rangle_{[\theta]}$ in its four arguments, relation (43) is established for the operators $\hat{R}_m$ (46) and R_m (48) too. From the proof of Theorem 3, recall that $C_{\hat{e}} = EC_eE^\top$ and $\hat{F}_{\hat{e}}^{-1} = EF_e^\dagger E^\top$. The relation (43) yields $\hat{R}_m(\hat{F}_{\hat{e}}^{-1}) = ER_m(F_e^\dagger)E^\top$. Substituting in the CRB gives:

$$EC_eE^\top \succeq E\left(F_e^\dagger - \frac{1}{3}\left(R_m(F_e^\dagger)F_e^\dagger + F_e^\dagger R_m(F_e^\dagger)\right)\right)E^\top,$$

where we used the fact that $R_m(M)P_e = P_e R_m(M) = R_m(M)$, which is easily established from (43). Lemma 1 applies and concludes the proof, since $P_e F_e^\dagger P_e = F_e^\dagger$. ■

V. Example

We give a simple example inspired from sensor network localization, to illustrate how both theorems for submanifolds and quotient manifolds can apply to the same problem. The parameter spaces are kept simple on purpose. In an upcoming paper, a more sophisticated setup will be considered where the objects to estimate are rotations, leading to curved parameter spaces [12].

Let $\theta = (\theta_1, \ldots, \theta_N)$ be a vector of N unknown but deterministic points in $\mathbb{R}^n$. Those can be thought of as positions, states, opinions, etc. of N agents. Let us consider an undirected graph on N nodes with edge set Ω, such that for each edge $\{i,j\} \in \Omega$ we have a noisy measurement of the relative state $v_{ij} = \theta_j - \theta_i + n_{ij}$, where the $n_{ij} \sim \mathcal{N}(0, \Sigma)$ are i.i.d. normally distributed noise vectors. By symmetry, $v_{ij} = -v_{ji}$, so $n_{ij} = -n_{ji}$. While it is important to assume independence of noise on distinct edges to keep the derivation simple, it is easy to relax the assumption that they have identical distributions. We assume identical distributions to keep notations simple.

The task is to estimate the θ_i's from the v_{ij}'s, thus $\mathcal{P} = (\mathbb{R}^n)^N$, and we set out to derive Cramér-Rao bounds for this problem. An alternative way of obtaining this result can be found in [13]. Decentralized algorithms to execute this synchronization can be found there and in [14].

The log-likelihood function $L : \mathcal{P} \to \mathbb{R}$ reads, with $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_N)$ and $V_i = \{j : \{i,j\} \in \Omega\}$ the set of neighbors of node i and dropping additive constants:

$$L(\hat{\theta}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j \in V_i} -\frac{1}{2} (v_{ij} - \hat{\theta}_j + \hat{\theta}_i)^\top \Sigma^{-1} (v_{ij} - \hat{\theta}_j + \hat{\theta}_i).$$

In order to compute the Fisher information matrix for this problem, we need to pick an orthonormal basis of $\mathrm{T}_\theta \mathcal{P} \equiv \mathcal{P}$. We choose the basis such that the first n vectors correspond to the canonical basis for the first copy of $\mathbb{R}^n$ in $\mathcal{P}$, the next n vectors correspond to the canonical basis for the second copy of $\mathbb{R}^n$ in $\mathcal{P}$, etc., totaling nN orthonormal basis vectors. The gradient of $L(\hat{\theta})$ w.r.t. $\hat{\theta}_i$ in this basis is the following vector in $\mathbb{R}^n$:

$$\mathrm{grad}_i L(\hat{\theta}) = \sum_{j \in V_i} \Sigma^{-1} (v_{ij} - \hat{\theta}_j + \hat{\theta}_i). \tag{50}$$

Hence, $\mathrm{grad}_i L(\theta) = \sum_{j \in V_i} \Sigma^{-1} n_{ij}$. The FIM F (3) is formed of $N \times N$ blocks of size $n \times n$. Due to independence of the n_{ij}'s and $n_{ij} = -n_{ji}$,

$$\mathbb{E}\left\{(\Sigma^{-1}n_{ij})(\Sigma^{-1}n_{k\ell})^\top\right\} = \Sigma^{-1}\mathbb{E}\left\{n_{ij}n_{k\ell}^\top\right\}\Sigma^{-1} = \begin{cases} \Sigma^{-1} & \text{if } (i,j) = (k,\ell), \\ -\Sigma^{-1} & \text{if } (i,j) = (\ell,k), \\ 0 & \text{otherwise.} \end{cases} \tag{51}$$

Hence, the $(i,j)^{\text{th}}$ block of the FIM is given by:

$$F_{ij} = \mathbb{E}\left\{\mathrm{grad}_i L(\theta) \cdot \mathrm{grad}_j L(\theta)^\top\right\} = \begin{cases} |V_i|\Sigma^{-1} & \text{if } i = j, \\ -\Sigma^{-1} & \text{if } \{i,j\} \in \Omega, \\ 0 & \text{otherwise.} \end{cases} \tag{52}$$

The structure of the graph Laplacian is apparent. Let $D = \mathrm{diag}(|V_1|, \ldots, |V_N|)$ be the degree matrix and let A be the adjacency matrix of the measurement graph. The Laplacian $\mathcal{L} = D - A$ is tied to the FIM via:

$$F = \mathcal{L} \otimes \Sigma^{-1}, \tag{53}$$

where $\otimes$ denotes the Kronecker product.

Of course, since we only have relative measurements, we can only hope to recover the θ_i's up to a global translation. And indeed, for every translation vector $t \in \mathbb{R}^n$, we have $L(\hat{\theta}) = L(\hat{\theta} + t)$, where $\hat{\theta} + t \triangleq (\hat{\theta}_1 + t, \ldots, \hat{\theta}_N + t)$. That is, all $\hat{\theta} + t$ induce the same distribution of the measurements v_{ij}, and are thus indistinguishable. This is the root of the rank deficiency of the FIM. Surely, if the graph is connected, the all-ones vector $\mathbb{1}_N$ forms a basis of $\ker \mathcal{L}$. Consequently, $\ker F$ consists of all vectors of the form $\mathbb{1}_N \otimes t$, with arbitrary $t \in \mathbb{R}^n$. Naturally, these correspond to global translations by t.

To resolve this ambiguity, we can either add constraints, most naturally in the form of anchors, or work on the quotient space.

a) With anchors: Let us consider $A \subset \{1, \ldots, N\}$, $A \neq \emptyset$, such that all θ_i with $i \in A$ are known; these are anchors. The resulting parameter space $\hat{\mathcal{P}} = \{\hat{\theta} \in \mathcal{P} : \hat{\theta}_i = \theta_i \ \forall i \in A\}$ is a submanifold of $\mathcal{P}$. The orthogonal projector from $\mathrm{T}_\theta\mathcal{P}$ to $\mathrm{T}_\theta\hat{\mathcal{P}}$ simply sets all components of a tangent vector corresponding to anchored nodes to zero. Formally, $P = I_A \otimes I_n$, where I_A is a diagonal matrix of size N whose i^{th} diagonal entry is 1 if $i \notin A$ and 0 otherwise. It follows that $PFP = I_A\mathcal{L}I_A \otimes \Sigma^{-1} = \mathcal{L}_A \otimes \Sigma^{-1}$, with the obvious definition for $\mathcal{L}_A$: the Laplacian with rows and columns corresponding to anchored nodes forced to zero. $\hat{\mathcal{P}}$ is a flat Euclidean space, hence its curvature tensor vanishes identically. Theorem 2 yields the anchored CRB for the covariance matrix C of an unbiased estimator on $\hat{\mathcal{P}}$:

$$\mathbb{E}\left\{(\hat{\theta}-\theta)(\hat{\theta}-\theta)^\top\right\} \triangleq C \succeq \mathcal{L}_A^\dagger \otimes \Sigma. \tag{54}$$

We used the fact that Kronecker product and pseudoinversion commute, see Proposition 6 in the appendix. This bound is easily interpreted in terms of individual nodes. Indeed, by definition, inequality (54) means that for all $x \in \mathbb{R}^{nN}$, $x^\top C x \geq x^\top(\mathcal{L}_A^\dagger \otimes \Sigma)x$. In particular, setting $x = e_i \otimes e_k$ with e_i the i^{th} canonical basis vector of $\mathbb{R}^N$ and e_k the k^{th} canonical basis vector of $\mathbb{R}^n$, we have:

$$\mathbb{E}\left\{(\hat{\theta}_i-\theta_i)_k^2\right\} \geq (\mathcal{L}_A^\dagger)_{ii} \cdot \Sigma_{kk}. \tag{55}$$

Summing over $k = 1 \ldots n$, this translates into a lower bound on the variance for estimating the state of node i:

$$\mathbb{E}\left\{\|\hat{\theta}_i-\theta_i\|^2\right\} \geq (\mathcal{L}_A^\dagger)_{ii} \cdot \operatorname{trace}(\Sigma). \tag{56}$$

This puts forward the importance of the diagonal of $\mathcal{L}_A^\dagger$. Taking traces on both sides of (54), we obtain an inequality for the total variance:

$$\mathbb{E}\left\{\operatorname{dist}^2(\hat{\theta},\theta)\right\} \triangleq \mathbb{E}\Big\{\sum_{i=1}^N \|\hat{\theta}_i-\theta_i\|^2\Big\} \geq \operatorname{trace}(\mathcal{L}_A^\dagger)\operatorname{trace}(\Sigma).$$

Notice that it would have been simple to pick a new basis for $\hat{\mathcal{P}}$, but this would have required a renumbering of the rows and columns of the matrices appearing in the CRB, which is slightly annoying. If the ambiguities are fixed not by adding anchors but, more generally, by adding one or more (for example) linear constraints of the form $a_1\theta_1 + \cdots + a_N\theta_N = b$, it becomes less obvious how to pick a meaningful basis for $\hat{\mathcal{P}}$ without breaking symmetry. In comparison, the projection method used here will apply gracefully, preserving symmetry and row/column ordering in the CRB matrices.

b) Without anchors: If there are no anchors, perhaps because there is no meaningful reference to begin with, we work on the quotient space $\hat{\mathcal{P}} = \mathcal{P}/\sim$, where $\theta \sim \theta'$ iff there exists a translation vector $t \in \mathbb{R}^n$ such that $\theta = \theta' + t$. The distance between the equivalence classes $[\theta]$ and $[\theta']$ on $\hat{\mathcal{P}}$ is the distance between their best aligned members, that is:

$$\operatorname{dist}^2([\theta],[\theta']) = \min_{t\in\mathbb{R}^n} \sum_{i=1}^N \|\theta_i + t - \theta_i'\|^2. \tag{57}$$

The optimal t is easily seen to be $t = \frac{1}{N}\sum_{i=1}^N \theta_i' - \theta_i$, which amounts to aligning the centers of mass of θ and θ'. Consequently, if we denote by θ_{c} the centered version of θ—i.e., θ translated such that its center of mass is at the origin—we find that:

$$\mathrm{dist}^2([\theta],[\theta']) = \mathrm{dist}^2(\theta_\mathrm{c},\theta'_\mathrm{c}) = \sum_{i=1}^N \|\theta_{\mathrm{c},i} - \theta'_{\mathrm{c},i}\|^2. \tag{58}$$

It follows that $\mathrm{dist}^2([\theta],[\theta']) = \mathrm{dist}^2(\theta_\mathrm{c},\theta'_\mathrm{c})$, hence the mapping $[\theta] \mapsto \theta_\mathrm{c}$ is an isometry between $\hat{\mathcal{P}}$ and the Euclidean space $\mathcal{P}$. We thus conclude that $\hat{\mathcal{P}}$ is a flat manifold and that its curvature tensor vanishes identically [10, Chap. 7]. Theorem 3 and Proposition 6 then yield:

$$\mathbb{E}\left\{(\hat{\theta}_\mathrm{c} - \theta_\mathrm{c})(\hat{\theta}_\mathrm{c} - \theta_\mathrm{c})^\top\right\} \triangleq C \succeq \mathcal{L}^\dagger \otimes \Sigma, \text{ and} \tag{59}$$

$$\mathbb{E}\Big\{\sum_{i=1}^N \|\hat{\theta}_{\mathrm{c},i} - \theta_{\mathrm{c},i}\|^2\Big\} \geq \mathrm{trace}(\mathcal{L}^\dagger)\mathrm{trace}(\Sigma). \tag{60}$$

We now interpret the CRB (59). Because of the ambiguity in the anchor-free scenario, it does not make much sense to ask what the variance for estimating a specific state is going to be. Rather, one should establish bounds for the variance on estimating the relative state between two nodes, i and j. Let $x = (e_i - e_j) \otimes e_k$ with e_i, e_j the i^th and j^th canonical basis vectors of $\mathbb{R}^N$ and e_k the k^th canonical basis vector of $\mathbb{R}^n$. Notice that x is a horizontal vector (its components sum to zero). Applying $x^\top \cdot x$ on both sides of (59) yields:

$$\mathbb{E}\left\{\left((\hat{\theta}_i - \hat{\theta}_j) - (\theta_i - \theta_j)\right)_k^2\right\} \geq (e_i - e_j)^\top \mathcal{L}^\dagger (e_i - e_j) \cdot \Sigma_{kk}.$$

Notice that there is no need to center $\hat{\theta}$ nor θ anymore, since the quantities involved are relative states. Summing over $k = 1 \ldots n$ gives a lower-bound on the variance for estimating the relative state between node i and node j:

$$\mathbb{E}\left\{\left\|(\hat{\theta}_i - \hat{\theta}_j) - (\theta_i - \theta_j)\right\|^2\right\} \geq (e_i - e_j)^\top \mathcal{L}^\dagger (e_i - e_j) \cdot \mathrm{trace}(\Sigma). \tag{61}$$

A nice interpretation is now possible. Indeed, the quantity $(e_i - e_j)^\top \mathcal{L}^\dagger (e_i - e_j)$ is well-known to correspond to the Euclidean commute time distance (ECTD) between nodes i and j [15]. It is small if many short paths connect the two nodes and if those paths have edges with large weights which, in our case, means measurements of high quality. Furthermore, the authors of [15] show how one can produce an embedding of the nodes in, say, the plane such that two nodes are close-by if the ECTD separating them is small. This is done via a projection akin to PCA and is an interesting visualization tool as it leads to a plot of the graph such that easily synchronizable nodes are clustered together.

Notice that the bound without anchors has a very different interpretation than that of the bound one obtains by artificially fixing an arbitrary node. This goes to showing that acknowledging the quotient nature of the parameter space does make a practical difference. Notice also that, since we did not need to switch to a different basis to obtain the bounds, regardless of which anchors we did or did not choose, it is always the same rows and columns of the matrices in the CRB's that refer to a specific node, which is rather convenient.

The maximum-likelihood estimator in the absence of anchors is easily obtained as the minimum-norm solution to the problem $\max L(\hat{\theta})$ (which is concave, quadratic). This estimator is centered and we state without proof that it is efficient, i.e., its covariance is exactly $\mathcal{L}^\dagger \otimes \Sigma$. In the anchored case, the maximum-likelihood estimator is best obtained via quadratic programming.

For the sake of simplicity, we considered a connected graph. In general, the graph might be disconnected, and there would then be more ambiguity. It is obvious that, in general, there is an $\mathbb{R}^n$ ambiguity for each connected component that does not include an anchor. The CRB's presented here can easily be derived to take care of this more general situation: one simply needs to redefine the equivalence relation $\sim$ accordingly. This in turn leads to a new quotient space with an appropriate notion of distance and covariance. The theorems established in this paper apply seamlessly to this more general scenario.

VI. Conclusions

We proposed four theorems that are meant to ease the use of the intrinsic Cramér-Rao bounds developed in [3] when the actual parameter space is a Riemannian submanifold or a Riemannian quotient manifold of a (usually more natural) parent space. We showed on a simple example how these theorems easily provide meaningful bounds for estimation problems with indeterminacies, whether these are dealt with by including prior knowledge or by acknowledging the quotient nature of the parameter space. We also observed on this same example that fixing indeterminacies by adding constraints results in a different CRB than if the quotient nature is acknowledged.

As previously observed in [2] and [13], the estimation problem in our example can be seen as a synchronization problem on the Lie group of translations, $\mathbb{R}^n$. In an upcoming paper, we will use the present tools to derive CRB's for synchronization on the Lie group of rotations for a wide family of noise models [12]. This will require the present results in all their generality, as curvature terms will not vanish anymore.

Acknowledgments

NB gratefully acknowledges J. Xavier and V. Barroso for figures 2–3 as well as P.-A. Absil and A. Singer for their useful comments about the original manuscript. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity

Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. NB is an FNRS research fellow (Aspirant). The scientific responsibility rests with its authors.

Appendix

Proposition 6 (pseudo-inverse and Kronecker product commute)**.** *Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ be any two real matrices. Then, $(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$, where $\otimes$ denotes the Kronecker product and $\dagger$ denotes the pseudo-inverse.*

Proof: Let r_A and r_B denote the ranks of A and B. The compact SVDs of A and B take the form $A = U_A \Sigma_A V_A^\top$ and $B = U_B \Sigma_B V_B^\top$, with $\Sigma_A \in \mathbb{R}^{r_A\times r_A}, \Sigma_B \in \mathbb{R}^{r_B\times r_B}$ full rank square matrices and $U_A^\top U_A = V_A^\top V_A = I_{r_A}$ and $U_B^\top U_B = V_B^\top V_B = I_{r_B}$. We will use three properties of the Kronecker product: for any real matrices W, X, Y, Z of suitable sizes and any invertible matrices N, M, we have (a) $(WX) \otimes (YZ) = (WY) \otimes (XZ)$, (b) $(A \otimes B)^\top = A^\top \otimes B^\top$, and (c) $(N \otimes M)^{-1} = N^{-1} \otimes M^{-1}$.

Recall that the pseudo-inverse of a matrix is given explicitly in terms of its compact SVD $X = U\Sigma V^\top$ by $X^\dagger = V\Sigma^{-1}U^\top$, such that

$$A^\dagger \otimes B^\dagger = (V_A \Sigma_A^{-1} U_A^\top) \otimes (V_B \Sigma_B^{-1} U_B^\top) \tag{62}$$

$$= (V_A \otimes V_B)(\Sigma_A^{-1} \otimes \Sigma_B^{-1})(U_A \otimes U_B)^\top, \tag{63}$$

where to reach the right-hand side we used properties (a–b). On the other hand, by properties (a–b) once again, we see that

$$A \otimes B = (U_A \otimes U_B)(\Sigma_A \otimes \Sigma_B)(V_A \otimes V_B)^\top. \tag{64}$$

The latter is a compact SVD for $A \otimes B$, hence the definition of pseudo-inverse implies that

$$(A \otimes B)^\dagger = (V_A \otimes V_B)(\Sigma_A \otimes \Sigma_B)^{-1}(U_A \otimes U_B)^\top. \tag{65}$$

Applying property (c) to $(\Sigma_A \otimes \Sigma_B)^{-1} = \Sigma_A^{-1} \otimes \Sigma_B^{-1}$ concludes the proof. ■

References

[1] R. Tron and R. Vidal, "Distributed image-based 3D localization of camera sensor networks," in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*. IEEE, 2009, pp. 901–908.

[2] A. Singer, "Angular synchronization by eigenvectors and semidefinite programming," *Applied and Computational Harmonic Analysis*, 2010.

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

[4] C. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, no. 3, pp. 81–91, 1945.

[5] J. Xavier and V. Barroso, "Intrinsic Variance Lower Bound (IVLB): an extension of the Cramér-Rao bound to Riemannian manifolds," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 5. IEEE, 2005, pp. 1033–1036.

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

[7] J. Xavier and V. Barroso, "The Riemannian geometry of certain parameter estimation problems with singular Fisher information matrices," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 2. IEEE, 2004, pp. 1021–1024.

[8] B. O'Neill, *Semi-Riemannian geometry: with applications to relativity*. Academic Pr, 1983, vol. 103.

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

[10] J. Lee, *Riemannian manifolds: An introduction to curvature*, ser. Graduate Texts in Mathematics. Springer, 1997, vol. 176.

[11] I. Chavel, *Riemannian geometry: a modern introduction*. Cambridge Univ Pr, 2006, vol. 98.

[12] N. Boumal, A. Singer, P.-A. Absil, and V. D. Blondel, "Cramér-Rao bounds for synchronization of rotations," Pôle de mathématiques appliquées, Université catholique de Louvain, Louvain-la-Neuve, Belgium, Tech. Rep. xxxx, July 2012.

[13] S. Howard, D. Cochran, W. Moran, and F. Cohen, "Estimation and registration on graphs," *Arxiv preprint arXiv:1010.2983*, 2010.

[14] W. Russell, D. Klein, and J. Hespanha, "Optimal estimation on the graph cycle space," *Signal Processing, IEEE Transactions on*, vol. 59, no. 6, pp. 2834–2846, 2011.

[15] M. Saerens, F. Fouss, L. Yen, and P. Dupont, "The principal components analysis of a graph, and its relationships to spectral clustering," *Machine Learning: ECML 2004*, pp. 371–383, 2004.