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

Nicolas Boumal

Department of Mathematical Engineering, 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 $\bar{\mathcal{P}}$ is a Riemannian submanifold or a Riemannian quotient manifold of a parent space $\mathcal{P}$, that is, estimation problems on manifolds 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 $\bar{\mathcal{P}}$. 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 two examples, we show how the CRB's for synchronization problems (including a type of sensor network localization problem) differ in the presence or absence of anchors, leading to bounds for 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 (loosely, nonlinear spaces with a smooth structure such as the sphere for example). In such problems, one would like to estimate a deterministic but unknown parameter θ belonging to a manifold $\mathcal{P}$, given a measurement y belonging to a probability space $\mathcal{M}$. The measurement y is a random variable whose probability density function is shaped by θ. It is because y is distributed differently for different θ's that sampling (observing) y reveals information about θ. We assume some familiarity of the reader with differential geometry. Appendix A briefly reviews the tools needed in this paper, fixes notation and provides references to classical introductory textbooks.

Estimation problems on manifolds 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 relate the covariance matrix of 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 the many references therein. We focus on bounds for unbiased estimators.

More formally, let $\mathcal{P}$ be a Riemannian manifold and $\mathcal{M}$ be a probability space, i.e., a measurable space with nonnegative 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

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

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. y):

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

where $\mathrm{T}_\theta\mathcal{P}$ is the tangent space to $\mathcal{P}$ at θ and $\mathrm{D}L(\theta)[u]$ denotes the directional derivative of L at θ along the tangent vector u (see Appendix A). 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 consider estimation problems such that $\mathbf{F}$ is *not necessarily* positive definite. Singularity of $\mathbf{F}$ 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 $\bar{\mathcal{P}} \subset \mathcal{P}$, 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 the sphere $\bar{\mathcal{P}} = \mathbb{S}^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$ 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 $\bar{\mathcal{P}} = \mathcal{P}/\sim$, where $\sim$ is an equivalence relation on $\mathcal{P}$ stating that $\theta, \theta' \in \mathcal{P}$ are equivalent if they give rise to the same distribution of the measurements. 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 direct way to address these issues would be to work on the smaller space $\bar{\mathcal{P}}$ directly, writing down Fisher information and covariance with respect to bases of the tangent spaces to $\bar{\mathcal{P}}$, 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 $\bar{\mathcal{P}}$ to begin with. Furthermore, when $\bar{\mathcal{P}}$ 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. The present work is based on [3] and derives 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 $\mathbb{R}^d$, has been studied extensively [6], [7], [8]. Notably, in [8], the authors describe $\bar{\mathcal{P}}$ 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 more generally consider Riemannian submanifolds of any Riemannian manifold $\mathcal{P}$. Furthermore, for the simple versions of the CRB's, only an orthogonal projector from the tangent spaces of $\mathcal{P}$ to those of $\bar{\mathcal{P}}$ are required. More importantly, the covariance matrix in the proposed bounds is expressed in terms of the Riemannian, or geodesic, distance on $\bar{\mathcal{P}}$, which may be more natural for a number of applications.

The case of CRB's for estimation problems with singular FIM has also been investigated extensively [7], [9], [10]. The classical remedy is to use the Moore-Penrose generalized inverse, hereafter referred to as the pseudoinverse, of the FIM instead of the inverse in the CRB. When the singularity is due to indeterminacies (a notion we make precise in Section III), Xavier and Barroso [10] showed a nice interpretation of the role of the pseudoinverse by recasting the estimation problem on a Riemannian quotient manifold $\bar{\mathcal{P}}$. In the latter reference, 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 possible curvature of $\bar{\mathcal{P}}$ is captured through a single number: an upperbound on the sectional curvatures of $\bar{\mathcal{P}}$. In comparison, since the present results are based on [3], the proposed bounds concern the whole covariance matrix (the trace of which coincides with the variance). The pseudoinverse of the FIM appears naturally through the same manipulations as in [10]. The additional curvature terms in the CRB (Section IV) take the whole Riemannian curvature tensor into account. This is especially useful when $\bar{\mathcal{P}}$ 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, i.e., less restrictive—hence less informative—because it has to assume maximum curvature in all directions. In comparison, the bounds derived here based on [3] are able to capture complex curvature structures if need be.

A lot of work has been accomplished to gain other types of geometric insight into the Cramér-Rao bounds. Scharf and McWhorter [11] study the CRB for multi-parameter estimation problems where the parameters are partitioned into two groups. Their study shows how the angles between the subspaces spanned by the sensitivity vectors of the two groups are related to the CRB. Another trend of geometric interpretation comes from the field of information geometry, where one endows the parameter space with Fisher information as a metric, turning the parameter space (typically, $\mathbb{R}^n$) into a Riemannian manifold [12]. The metric "bends" space, so that the larger the Riemannian distance between two parameters, the easier it is for an estimator to distinguish them.

The CRB's presented in this communication hold *at large SNR*. The origin of this provision is double. Firstly, the definition of covariance on a manifold uses the logarithmic map on that manifold, which is only locally well-defined. It is thus necessary to require the noise level to be low enough so that the estimator $\hat{\theta}$ of a parameter θ will, with high probability, belong to a neighborhood of θ where the logarithm is well-defined. This can be relevant even on flat manifolds such as the circle $\mathrm{SO}(2)$ for example, which is compact. Secondly, on curved manifolds, the proof of the main theorem in [3] relies on truncated Taylor expansions. Those are legitimate only at large enough SNR so that typical errors are small compared to the scale at which curvature becomes a dominant feature.

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 a $d \times d$ symmetric matrix defined by:

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

where L is the log-likelihood function (1). 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 that may appear due to the possible curvature of $\bar{\mathcal{P}}$. This will result in simple statements (Theorems 2–3). These are practically useful because the curvature terms are often negligible at large SNR [3]. Then, we will establish the CRB's including curvature terms (Section IV). Finally, we will illustrate the usage of these theorems through two examples (Section V).

II. RIEMANNIAN SUBMANIFOLDS

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

Let $\hat{\theta}$ be any unbiased estimator for the estimation problem, that is, $\hat{\theta} : \mathcal{M} \to \bar{\mathcal{P}}$ maps every possible realization of the measurement y to a parameter $\hat{\theta}(y)$ and has zero bias:

$$\forall \theta \in \mathcal{P}, \quad b(\theta) = \mathbb{E}\left\{\mathrm{Log}_\theta(\hat{\theta}(y))\right\} = 0, \tag{4}$$

where $\mathrm{Log}_\theta : \bar{\mathcal{P}} \to \mathrm{T}_\theta\bar{\mathcal{P}}$ is the logarithmic map at θ on $\bar{\mathcal{P}}$ (see Appendix A). For example, on a Euclidean space, $\mathrm{Log}_\theta(\hat{\theta}(y)) = \hat{\theta}(y) - \theta$. For conciseness, we often write $\hat{\theta}$ to mean $\hat{\theta}(y)$. The covariance matrix of $\hat{\theta}$ w.r.t. the basis e is defined as:

$$(C_e)_{ij} = \mathbb{E}\left\{\left\langle \mathrm{Log}_\theta(\hat{\theta}), e_i \right\rangle_\theta \cdot \left\langle \mathrm{Log}_\theta(\hat{\theta}), e_j \right\rangle_\theta\right\}, \tag{5}$$

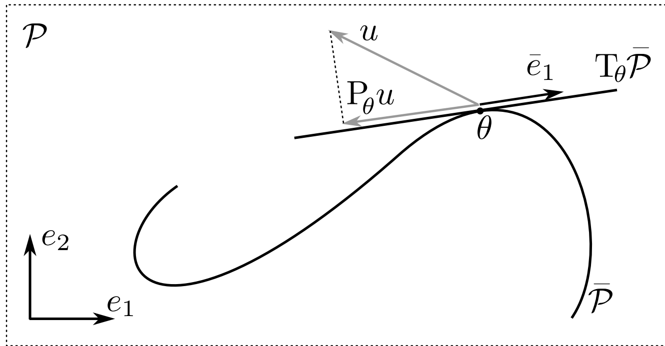

Fig. 1. $\bar{\mathcal{P}}$ is a Riemannian submanifold of $\mathcal{P}$. We consider estimation problems for which the parameter to estimate is θ, a point of $\bar{\mathcal{P}}$. In this drawing, for simplicity, we chose $\mathcal{P} = \mathbb{R}^2$. The vectors $e = (e_1, e_2)$ form an orthonormal basis of $\mathrm{T}_\theta\mathcal{P} \equiv \mathcal{P}$, while $\bar{e} = (\bar{e}_1)$ is an orthonormal basis of the tangent space $\mathrm{T}_\theta\bar{\mathcal{P}}$. The operator P_θ projects vectors of $\mathrm{T}_\theta\mathcal{P}$ orthogonally onto $\mathrm{T}_\theta\bar{\mathcal{P}}$. 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 $\bar{e}$ for each point θ.

where, as always in this paper, the expectation is taken w.r.t. the measurements $y \sim f(y;\theta)$. The goal is to link C_e and F_e through a matrix inequality.

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

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

$$\begin{aligned}
(C_{\bar{e}})_{ij} &= \mathbb{E}\left\{\left\langle \mathrm{Log}_\theta(\hat{\theta}), \bar{e}_i\right\rangle_\theta \cdot \left\langle \mathrm{Log}_\theta(\hat{\theta}), \bar{e}_j\right\rangle_\theta\right\}, \\
(\bar{F}_{\bar{e}})_{ij} &= \bar{\mathbf{F}}[\bar{e}_i, \bar{e}_j] = \mathbf{F}[\bar{e}_i, \bar{e}_j], \\
C_{\bar{e}} &\succeq \bar{F}_{\bar{e}}^{-1} + \text{ curvature terms.}
\end{aligned} \tag{6}$$

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 $\bar{e}$. This is what the next theorem achieves, right after we establish a technical lemma.

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

$$EAE^\top \succeq B \quad \Rightarrow \quad A \succeq E^\top BE. \tag{7}$$

Proof: Since $\mathbb{R}^d = \operatorname{im} E^\top \oplus \ker E$, for all $x \in \mathbb{R}^d$, there exist unique vectors $y \in \mathbb{R}^{\bar{d}}$ 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 $\bar{\mathcal{P}}$, at large SNR, the $d \times d$ covariance matrix C_e* (5) *and the $d \times d$ Fisher information matrix F_e* (3) *obey the matrix inequality (assuming* $\operatorname{rank}(P_e F_e P_e) = \bar{d}$*):*

$$C_e \succeq (P_e F_e P_e)^\dagger + \text{ curvature terms}, \tag{8}$$

where the $d \times d$ matrix $P_e = E^\top E$ is the orthogonal projector from $\mathrm{T}_\theta \mathcal{P}$ to $\mathrm{T}_\theta \bar{\mathcal{P}}$ 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 $\bar{F}_{\bar{e}}^{-1}$ with $d - \bar{d}$ additional zeroes. In particular, neglecting curvature terms:

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

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

$$C_e = E^\top C_{\bar{e}} E. \tag{9}$$

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

$$\bar{F}_{\bar{e}} = E F_e E^\top. \tag{10}$$

Notice that the assumption $\operatorname{rank}(P_e F_e P_e) = \bar{d}$ is equivalent to the assumption that $\bar{F}_{\bar{e}}$ is invertible. Then, substituting in (6), we find $E C_e E^\top \succeq (E F_e E^\top)^{-1}$. Since $\ker C_e = \ker(E^\top C_{\bar{e}} E) \supset \ker E$, Lemma 1 applies and it follows that (neglecting curvature terms):

$$C_e \succeq E^\top (E F_e E^\top)^{-1} E. \tag{11}$$

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

$$E^\top(EF_eE^\top)^{-1}E = (E^\top EF_eE^\top E)^\dagger. \tag{12}$$

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

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

$$(P_eF_eP_e)^\dagger = E^\top UDU^\top E = V \begin{pmatrix} D & \\ & 0 \end{pmatrix} V^\top, \tag{13}$$

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

$$\mathrm{trace}(C_e) = \mathrm{trace}(E^\top C_{\bar{e}}E) = \mathrm{trace}(C_{\bar{e}}) \geq \mathrm{trace}(\bar{F}_{\bar{e}}^{-1}) = \mathrm{trace}((P_eF_eP_e)^\dagger). \tag{14}$$

■

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 $\bar{\mathcal{P}}$:

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

Here is one way of interpreting the bound (8). Expand the random error vector $\mathrm{Log}_\theta(\hat{\theta}) = \sum_i x_i e_i$ with random coefficients x_i. From the definition, $(C_e)_{ii} = \mathbb{E}\left\{x_i^2\right\}$. Then, equation (8) implies $\mathbb{E}\left\{x_i^2\right\} \geq (P_eF_eP_e)_{ii}^\dagger$, which limits how well the i^{th} coordinate can be estimated. For example, when $\bar{\mathcal{P}}$ is Euclidean, $\mathrm{Log}_\theta(\hat{\theta}) = \hat{\theta} - \theta$ and $\mathbb{E}\left\{x_i^2\right\} = \mathbb{E}\left\{(\hat{\theta}_i - \theta_i)^2\right\}$.

Notice that it is not necessary to explicitly construct a basis $\bar{e}$ 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. This is fortunate since, because of the hairy ball theorem, it is impossible to define bases $\bar{e}$ 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:

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

The quotient space $\bar{\mathcal{P}} = \mathcal{P}/\sim$—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 [10], depict the concept of quotient manifold and of the related basic objects we introduce hereafter, namely submersions and horizontal/vertical spaces. See also Appendix A.

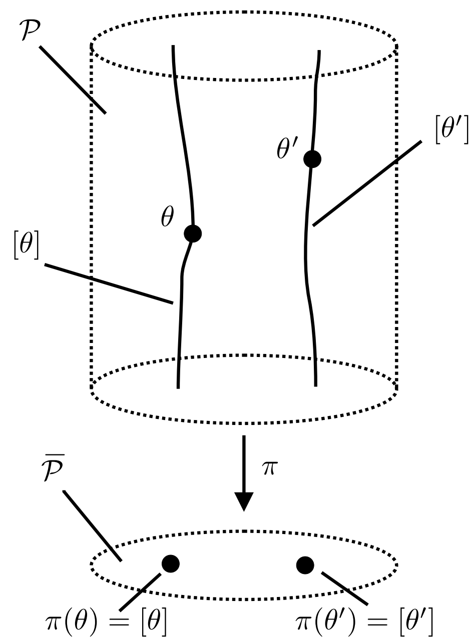

Fig. 2. The parameter space $\mathcal{P}$ is partitioned into equivalence classes, called fibers. The Riemannian submersion π maps each $\theta \in \mathcal{P}$ to its corresponding equivalence class $[\theta] \in \bar{\mathcal{P}}$. The space of equivalence classes is the quotient space $\bar{\mathcal{P}} = \mathcal{P}/\sim$, also a Riemannian manifold. *Figure courtesy of [10].*

We now consider the mapping π from $\mathcal{P}$ to $\bar{\mathcal{P}}$, which maps each parameter θ to its equivalence class $[\theta]$:

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

and concentrate on the case where π is a Riemannian submersion [13][14]. That is, $\bar{\mathcal{P}}$ is a Riemannian quotient manifold of $\mathcal{P}$. In particular, $[\theta]$ is a Riemannian submanifold of $\mathcal{P}$ (a *fiber*). The log-likelihood function $\bar{L} : \bar{\mathcal{P}} \to \mathbb{R}$ is well-defined by $\bar{L}([\theta]) \triangleq L(\theta)$.

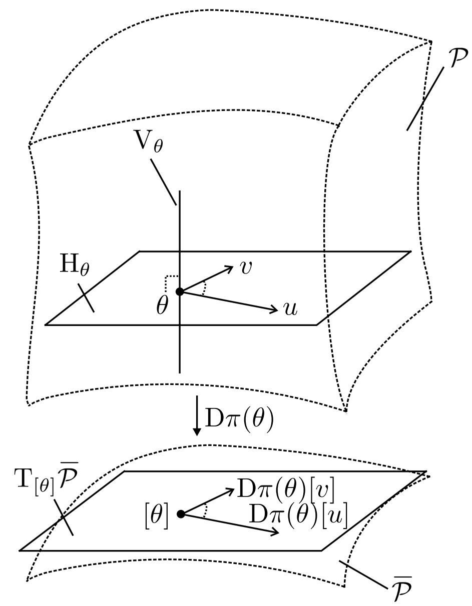

Fig. 3. Each fiber $\pi(\theta) = [\theta]$ is a Riemannian submanifold of $\mathcal{P}$. The tangent space to a fiber at θ is the vertical space V_θ. The orthogonal complement of V_θ in $\mathrm{T}_\theta\mathcal{P}$ is the horizontal space H_θ. The differential of π, noted $\mathrm{D}\pi(\theta)$, is an isometry between H_θ and the abstract tangent space $\mathrm{T}_{[\theta]}\bar{\mathcal{P}}$. This makes it convenient to represent abstract tangent vectors to $\bar{\mathcal{P}}$ as horizontal vectors. *Figure courtesy of [10].*

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 $\mathrm{D}\pi(\theta) : \mathrm{T}_\theta\mathcal{P} \to \mathrm{T}_{[\theta]}\bar{\mathcal{P}}$ of a Riemannian submersion induces a metric on the abstract tangent space $\mathrm{T}_{[\theta]}\bar{\mathcal{P}}$:

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

The definition of Riemannian submersion ensures that this is well-defined [14]. We mention two useful properties:

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

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

Let $[\hat{\theta}] : \mathcal{M} \to \bar{\mathcal{P}}$ be any unbiased estimator for the present 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 $\mathrm{Exp}_\theta(\xi) \in [\hat{\theta}]$. The exponential map is the inverse of the logarithmic map, see Appendix A. On a Euclidean space, $\mathrm{Exp}_\theta(\xi) = \theta + \xi$.

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

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

Since $\mathrm{T}_{[\theta]}\bar{\mathcal{P}}$ 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 $\bar{\mathcal{P}} = \mathcal{P}/\sim$* (15)*, at large SNR, the $d \times d$ covariance matrix C_e* (20) *and the $d \times d$ Fisher information matrix F_e* (3) *obey the matrix inequality (assuming $\mathrm{rank}(F_e) = \bar{d}$):*

$$C_e \succeq F_e^{\dagger} + \textit{ curvature terms}, \tag{22}$$

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

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

Proof: It is convenient to introduce the orthonormal basis of H_θ related to $\bar{e}$ as $\tilde{e} = (\tilde{e}_1, \ldots, \tilde{e}_{\bar{d}})$, with $\bar{e}_i = \mathrm{D}\pi(\theta)[\tilde{e}_i]$. The $\bar{d} \times d$ matrix E such that $E_{ij} = \langle \tilde{e}_i, e_j \rangle_\theta$ will prove useful. E is orthonormal: $EE^\top = I_{\bar{d}}$, 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 $\mathrm{D}\pi(\theta)[e_i] \in \mathrm{T}_{[\theta]}\bar{\mathcal{P}}$ expands in the basis $\bar{e}$ as $\mathrm{D}\pi(\theta)[e_i] = \sum_j E_{ji}\bar{e}_j$. Indeed,

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

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

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

Since $EE^\top = I_{\bar{d}}$, it also holds that $C_{\bar{e}} = EC_eE^\top$.

We now similarly link F_e and $\bar{F}_{\bar{e}}$. 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).

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

In matrix form,

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

This highlights the fact that $\ker F_e = \ker E$ (since $\bar{F}_{\bar{e}}$ is invertible), which makes sense since $\ker E$ is the vertical space V_θ (more precisely, it is the space of coordinate vectors of vertical vectors w.r.t. the basis e). Again, by orthonormality of E, it also holds that $\bar{F}_{\bar{e}} = EF_eE^\top$. Combining these rules, it follows that:

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

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

$$C_e \succeq F_e^\dagger + \text{ curvature terms,} \tag{30}$$

since

$$E^\top (E F_e E^\top)^{-1} E = (E^\top E F_e E^\top E)^\dagger = F_e^\dagger. \tag{31}$$

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

Again, there is no need to construct bases $\tilde{e}$ or $\bar{e}$ in order to use Theorem 3. Notice that it still holds that $\mathrm{trace}(C_e) = \mathrm{trace}(C_{\bar{e}}) = \mathbb{E}\left\{\|\xi\|_\theta^2\right\} = \mathbb{E}\left\{\mathrm{dist}^2([\theta], [\hat{\theta}])\right\}$, where dist is the Riemannian distance on $\bar{\mathcal{P}}$, 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 $\bar{\mathcal{P}}$—for intuition on flatness and curvature, see Appendix A; for a reference, see [13], [15]. As noted in [3, p. 1615], "*the proof of the CRB in Euclidean spaces relies on the fact that* $\partial/\partial\theta(\hat{\theta}-\theta) = -I$. *However, for arbitrary Riemannian manifolds,* $\nabla \mathrm{Log}_\theta(\hat{\theta}) = -I$ + *second- and higher-order terms involving the manifold's sectional and Riemannian curvatures,*" where ∇ denotes the Riemannian connection on $\bar{\mathcal{P}}$, i.e., a notion of derivatives of vector fields on Riemannian manifolds. In a nutshell: the origin of curvature terms in intrinsic CRB's is the non-commutativity of derivatives of vector fields on curved spaces. As further noted [3, p. 1618], "*the significance of the curvature terms is an open question that depends on the specific application; however, as noted earlier, these terms become negligible for small errors and biases.*"

The curvature terms vanish if $\bar{\mathcal{P}}$ is flat, that is, if it is locally isometric to a Euclidean space. In such cases, theorems 2 and 3 suffice. When $\bar{\mathcal{P}}$ 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

$$\mathrm{dist}(\theta, \hat{\theta}) \ll \frac{1}{\sqrt{K_{\max}}}, \tag{32}$$

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

Condition (32) involves an upperbound on the sectional curvature of $\bar{\mathcal{P}}$. 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 in all product manifolds.

As an example, let us consider the problem of estimating $(\theta_1, \ldots, \theta_N) \in \bar{\mathcal{P}} = \mathbb{S}^2 \times \cdots \times \mathbb{S}^2$, the product of N spheres. $\bar{\mathcal{P}}$ has unit curvature along tangent 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 $\mathrm{dist}(\theta_i, \hat{\theta}_i)$ to be independent of i. Consequently, $\mathrm{dist}(\theta, \hat{\theta})$ grows as $\sqrt{N}$, whereas $K_{\max}$ remains constant. Hence, condition (32) 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 $\bar{\mathcal{P}}$ 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 $\bar{\mathcal{P}}$. Useful references to look up/compute this tensor are [13, Lemma 3.39, Cor. 3.58, Thm 7.47, Cor. 11.10][15][16].

A. Curvature terms for submanifolds

The random error vector $X \triangleq \mathrm{Log}_\theta(\hat{\theta})$ expands in the basis $\bar{e}$ as $X = \sum_i \bar{x}_i \bar{e}_i$, with $\bar{x}_1, \ldots, \bar{x}_{\bar{d}}$ random variables. Notice that $(C_{\bar{e}})_{ij} = \mathbb{E}\left\{\langle X, \bar{e}_i\rangle_\theta \langle X, \bar{e}_j\rangle_\theta\right\} = \mathbb{E}\left\{\bar{x}_i \bar{x}_j\right\}$. Let $\bar{R}$ be the Riemannian curvature tensor of $\bar{\mathcal{P}}$. The mapping $(u, v, w, z) \in (\mathrm{T}_\theta\bar{\mathcal{P}})^4 \mapsto \langle \bar{R}(u, v)w, z\rangle_\theta$ is linear in its four arguments [15]. Smith introduces the symmetric 2-form $\bar{\mathbf{R}}_{\mathbf{m}} : \mathrm{T}_\theta\bar{\mathcal{P}} \times \mathrm{T}_\theta\bar{\mathcal{P}} \to \mathbb{R}$ defined by [3, eq. (34)]:

$$\bar{\mathbf{R}}_{\mathbf{m}}[\bar{e}_i, \bar{e}_j] = \mathbb{E}\left\{\left\langle \bar{R}(X, \bar{e}_i)\bar{e}_j, X\right\rangle_\theta\right\} \tag{33}$$

$$= \mathbb{E}\left\{\sum_{k,\ell} \left\langle \bar{R}(\bar{e}_k, \bar{e}_i)\bar{e}_j, \bar{e}_\ell\right\rangle_\theta \bar{x}_k \bar{x}_\ell\right\} \tag{34}$$

$$= \sum_{k,\ell} \left\langle \bar{R}(\bar{e}_k, \bar{e}_i)\bar{e}_j, \bar{e}_\ell\right\rangle_\theta (C_{\bar{e}})_{k\ell}. \tag{35}$$

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

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

$$(\bar{R}_m(M))_{ij} = \sum_{k,\ell} \left\langle \bar{R}(\bar{e}_k, \bar{e}_i)\bar{e}_j, \bar{e}_\ell\right\rangle_\theta M_{k\ell}. \tag{36}$$

At large SNR, the CRB with curvature terms in [3, Cor. 2] reads

$$C_{\bar{e}} \succeq \bar{F}_{\bar{e}}^{-1} - \frac{1}{3}\left(\bar{R}_m(\bar{F}_{\bar{e}}^{-1})\bar{F}_{\bar{e}}^{-1} + \bar{F}_{\bar{e}}^{-1}\bar{R}_m(\bar{F}_{\bar{e}}^{-1})\right). \tag{37}$$

In order to provide an equivalent of (37) only referencing the basis e, we introduce the following symmetric 2-form on $\mathrm{T}_\theta\mathcal{P} \times \mathrm{T}_\theta\mathcal{P}$:

$$\mathbf{R_m}[e_i, e_j] \triangleq \bar{\mathbf{R}}_{\mathbf{m}}[P_\theta e_i, P_\theta e_j]. \tag{38}$$

Notice that, since $X \in \mathrm{T}_\theta\bar{\mathcal{P}}$, 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 $(C_e)_{ij} = \mathbb{E}\left\{x_i x_j\right\}$. It follows that:

$$\begin{aligned}\mathbf{R_m}[e_i, e_j] &= \mathbb{E}\left\{\left\langle \bar{R}(X, P_\theta e_i)P_\theta e_j, X\right\rangle_\theta\right\} &(39)\\ &= \sum_{k,\ell} \left\langle \bar{R}(P_\theta e_k, P_\theta e_i)P_\theta e_j, P_\theta e_\ell\right\rangle_\theta (C_e)_{k\ell}. &(40)\end{aligned}$$

From there, we introduce the following linear map:

$$\begin{aligned}&R_m : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d} : M \mapsto R_m(M), \text{ with}\\ &(R_m(M))_{ij} = \sum_{k,\ell} \left\langle \bar{R}(P_\theta e_k, P_\theta e_i)P_\theta e_j, P_\theta e_\ell\right\rangle_\theta M_{k\ell}. \qquad (41)\end{aligned}$$

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 $\mathbf{R_m}$:

$$4\mathbf{R_m}[e_i, e_j] = \mathbf{R_m}[e_i + e_j, e_i + e_j] - \mathbf{R_m}[e_i - e_j, e_i - e_j]. \tag{42}$$

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

Theorem 4 (CRB on submanifolds, with curvature)**.** *(Continued from Theorem 2). Including terms due to the possible curvature of $\bar{\mathcal{P}}$, at large SNR, the covariance matrix C_e* (5) *of any unbiased estimator $\hat{\theta} : \mathcal{M} \to \bar{\mathcal{P}}$ 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 $\mathrm{rank}(P_e F_e P_e) = \bar{d}$):*

$$C_e \succeq \tilde{F}_e^\dagger - \frac{1}{3}\left(R_m(\tilde{F}_e^\dagger)\tilde{F}_e^\dagger + \tilde{F}_e^\dagger R_m(\tilde{F}_e^\dagger)\right), \tag{43}$$

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 (41).

Proof: We start from the CRB w.r.t. the basis $\bar{e}$ (37):

$$C_{\bar{e}} \succeq \bar{F}_{\bar{e}}^{-1} - \frac{1}{3}\left(\bar{R}_m(\bar{F}_{\bar{e}}^{-1})\bar{F}_{\bar{e}}^{-1} + \bar{F}_{\bar{e}}^{-1}\bar{R}_m(\bar{F}_{\bar{e}}^{-1})\right). \tag{44}$$

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

$$\forall M \in \mathbb{R}^{d\times d}, \quad R_m(M) = E^\top \bar{R}_m(EME^\top)E. \tag{45}$$

From the proof of Theorem 2, recall that $C_{\bar{e}} = EC_eE^\top$ and $\bar{F}_{\bar{e}}^{-1} = E(P_eF_eP_e)^\dagger E^\top$. The relation (45) yields $\bar{R}_m(\bar{F}_{\bar{e}}^{-1}) = ER_m((P_eF_eP_e)^\dagger)E^\top$. 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 (45). 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 $\bar{e}$ as $X = \sum_i \bar{x}_i\bar{e}_i$, with $\bar{x}_1, \ldots, \bar{x}_{\bar{d}}$ random variables and $(C_{\bar{e}})_{ij} = \mathbb{E}\{\bar{x}_i\bar{x}_j\}$. Let $\bar{R}$ be the Riemannian curvature tensor of $\bar{\mathcal{P}}$. We consider $\bar{\mathbf{R}}_\mathbf{m} : \mathrm{T}_{[\theta]}\bar{\mathcal{P}} \times \mathrm{T}_{[\theta]}\bar{\mathcal{P}} \to \mathbb{R}$ defined by:

$$\bar{\mathbf{R}}_\mathbf{m}[\bar{e}_i, \bar{e}_j] = \mathbb{E}\left\{\langle \bar{R}(X, \bar{e}_i)\bar{e}_j, X\rangle_{[\theta]}\right\} \tag{46}$$

$$= \sum_{k,\ell} \langle \bar{R}(\bar{e}_k, \bar{e}_i)\bar{e}_j, \bar{e}_\ell\rangle_{[\theta]} (C_{\bar{e}})_{k\ell}. \tag{47}$$

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

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

$$(\bar{R}_m(M))_{ij} = \sum_{k,\ell} \langle \bar{R}(\bar{e}_k, \bar{e}_i)\bar{e}_j, \bar{e}_\ell\rangle_{[\theta]} M_{k\ell}. \tag{48}$$

Again, at large SNR, the CRB (37) holds. To express it only referencing the basis e, we introduce the following symmetric 2-form:

$$\mathbf{R}_\mathbf{m}[e_i, e_j] \triangleq \bar{\mathbf{R}}_\mathbf{m}\left[\mathrm{D}\pi(\theta)[e_i], \mathrm{D}\pi(\theta)[e_j]\right]. \tag{49}$$

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 $\xi = \sum_i x_ie_i$, we find $X = \sum_i x_i\mathrm{D}\pi(\theta)[e_i]$ with random variables $x_1, \ldots, x_d$ and $(C_e)_{ij} = \mathbb{E}\{x_ix_j\}$. It follows that:

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

From there, we introduce the following linear map:

$$R_m : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d} : M \mapsto R_m(M), \text{ with} \tag{50}$$

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

Theorem 5 (CRB on quotient manifolds, with curvature)**.** *(Continued from Theorem 3). Including terms due to the possible curvature of $\bar{\mathcal{P}}$, at large SNR, the covariance matrix C_e* (20) *of any unbiased estimator $\hat{\theta} : \mathcal{M} \to \bar{\mathcal{P}}$ 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 $\operatorname{rank}(F_e) = \bar{d}$):*

$$C_e \succeq F_e^\dagger - \frac{1}{3}\left(R_m(F_e^\dagger)F_e^\dagger + F_e^\dagger R_m(F_e^\dagger)\right), \tag{51}$$

where $R_m : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ is as defined by (50)*.*

Proof: The proof is very similar to that of Theorem 4. We start from the CRB w.r.t. the basis $\bar{e}$ (37). Expanding $\mathrm{D}\pi(\theta)[e_i] = \mathrm{D}\pi(\theta)[P\mathrm{H}_\theta e_i] = \sum_j \langle \tilde{e}_j, e_i \rangle \mathrm{D}\pi(\theta)[\tilde{e}_j] = \sum_j E_{ji}\bar{e}_j$ and exploiting linearity of $\langle \bar{R}(\cdot,\cdot)\cdot,\cdot\rangle_{[\theta]}$ in its four arguments, relation (45) is established for the operators $\bar{R}_m$ (48) and R_m (50) too. From the proof of Theorem 3, recall that $C_{\bar{e}} = EC_eE^\top$ and $\bar{F}_{\bar{e}}^{-1} = EF_e^\dagger E^\top$. The relation (45) yields $\bar{R}_m(\bar{F}_{\bar{e}}^{-1}) = ER_m(F_e^\dagger)E^\top$. 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 $R_m(M)P_e = P_eR_m(M) = R_m(M)$, which is easily established from (45). Lemma 1 applies and concludes the proof, since $P_eF_e^\dagger P_e = F_e^\dagger$. ■

V. Examples

We take a look at two examples of the family of synchronization problems [2]. In such problems, one considers a group $\mathcal{G}$ and a set of N group elements $g_1, \ldots, g_N \in \mathcal{G}$. The g_i's are to be estimated based on noisy measurements of group element ratios $g_ig_j^{-1}$. When $\mathcal{G}$ has a manifold structure, that is, when it is a Lie group, synchronization falls within the spectrum of estimation on manifolds. The first example is synchronization on the group of translations $\mathbb{R}^n$, which makes for a simple geometry and helps fix ideas. The second example is a simplified version of synchronization on $\mathrm{SO}(3)$, the group of rotations in $\mathbb{R}^3$. A more elaborate treatment of synchronization on $\mathrm{SO}(n)$ is given in [17]. Synchronization problems illustrate how both theorems for submanifolds and quotient manifolds can apply to the same setting, with rich interpretation.

A. Synchronization of translations

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$ 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 $\mathcal{E}$, such that for each edge $\{i, j\} \in \mathcal{E}$ we have a noisy measurement of the relative state $h_{ij} = \theta_j - \theta_i + n_{ij}$, where the $n_{ij} \sim \mathcal{N}(0, \Sigma)$ are i.i.d. normally distributed noise vectors. By symmetry, $h_{ij} = -h_{ji}$, 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 h_{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 [18]. Decentralized algorithms to execute this synchronization can be found there and in [19].

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

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

In order to compute the Fisher information matrix for this problem, we need to pick an orthonormal basis of $\mathrm{T}_{\boldsymbol{\theta}}\mathcal{P} \equiv \mathcal{P}$. 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 $L(\hat{\boldsymbol{\theta}})$ w.r.t. $\hat{\theta}_i$ in this basis is the following vector in $\mathbb{R}^n$:

$$\mathrm{grad}_i L(\hat{\boldsymbol{\theta}}) = \sum_{j \in V_i} \Sigma^{-1} (h_{ij} - \hat{\theta}_j + \hat{\theta}_i). \tag{52}$$

Hence, $\mathrm{grad}_i L(\boldsymbol{\theta}) = \sum_{j \in V_i} \Sigma^{-1} n_{ij}$. 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 $n_{ij} = -n_{ji}$,

$$\mathbb{E}\left\{(\Sigma^{-1} n_{ij})(\Sigma^{-1} n_{k\ell})^\top\right\} = \Sigma^{-1} \mathbb{E}\left\{n_{ij} n_{k\ell}^\top\right\} \Sigma^{-1} = \begin{cases} \Sigma^{-1} & \text{if } (i, j) = (k, \ell), \\ -\Sigma^{-1} & \text{if } (i, j) = (\ell, k), \\ 0 & \text{otherwise.} \end{cases} \tag{53}$$

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

$$F_{ij} = \mathbb{E}\left\{\mathrm{grad}_i L(\boldsymbol{\theta}) \cdot \mathrm{grad}_j L(\boldsymbol{\theta})^\top\right\} = \begin{cases} |V_i| \Sigma^{-1} & \text{if } i = j, \\ -\Sigma^{-1} & \text{if } \{i, j\} \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases} \tag{54}$$

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:

$$F = \mathcal{L} \otimes \Sigma^{-1}, \tag{55}$$

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 $L(\hat{\boldsymbol{\theta}}) = L(\hat{\boldsymbol{\theta}} + t)$, where $\hat{\boldsymbol{\theta}} + t \overset{\text{not.}}{=} (\hat{\theta}_1 + t, \ldots, \hat{\theta}_N + t)$. That is, all $\hat{\boldsymbol{\theta}} + t$ induce the same distribution of the measurements h_{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 $\bar{\mathcal{P}} = \{\hat{\boldsymbol{\theta}} \in \mathcal{P} : \hat{\theta}_i = \theta_i \ \forall i \in A\}$ is a submanifold of $\mathcal{P}$. The orthogonal projector from $\mathrm{T}_{\boldsymbol{\theta}}\mathcal{P}$ to $\mathrm{T}_{\boldsymbol{\theta}}\bar{\mathcal{P}}$ 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. $\bar{\mathcal{P}}$ is Euclidean, hence it is flat and its curvature tensor vanishes identically. Theorem 2 yields the anchored CRB for the covariance matrix C of an unbiased estimator on $\bar{\mathcal{P}}$:

$$\mathbb{E}\left\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top\right\} \triangleq C \succeq \mathcal{L}_A^\dagger \otimes \Sigma. \tag{56}$$

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 (56) 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:

$$\mathbb{E}\left\{(\hat{\theta}_i - \theta_i)_k^2\right\} \geq (\mathcal{L}_A^\dagger)_{ii} \cdot \Sigma_{kk}. \tag{57}$$

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

$$\mathbb{E}\left\{\|\hat{\theta}_i - \theta_i\|^2\right\} \geq (\mathcal{L}_A^\dagger)_{ii} \cdot \mathrm{trace}(\Sigma). \tag{58}$$

This puts forward the importance of the diagonal of $\mathcal{L}_A^\dagger$. Taking traces on both sides of (56), we obtain an inequality for the total variance:

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

Notice that it would have been simple to pick a new basis for $\mathrm{T}_{\boldsymbol{\theta}}\bar{\mathcal{P}}$, but this would have required a renumbering of the rows and columns of the matrices appearing in the CRB. 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 $\mathrm{T}_{\boldsymbol{\theta}}\bar{\mathcal{P}}$ 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 $\bar{\mathcal{P}} = \mathcal{P}/\sim$, where $\boldsymbol{\theta} \sim \boldsymbol{\theta}'$ iff there exists a translation vector $t \in \mathbb{R}^n$ such that $\boldsymbol{\theta} = \boldsymbol{\theta}' + t$. The distance between the equivalence classes $[\boldsymbol{\theta}]$ and $[\boldsymbol{\theta}']$ on $\bar{\mathcal{P}}$ is the distance between their best aligned members, that is:

$$\mathrm{dist}^2([\boldsymbol{\theta}], [\boldsymbol{\theta}']) = \min_{t\in\mathbb{R}^n} \sum_{i=1}^N \|\theta_i + t - \theta_i'\|^2. \tag{59}$$

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 $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. Consequently, if we denote by $\boldsymbol{\theta}_\mathrm{c}$ the centered version of $\boldsymbol{\theta}$—i.e., $\boldsymbol{\theta}$ translated such that its center of mass is at the origin—we find that:

$$\mathrm{dist}^2([\boldsymbol{\theta}], [\boldsymbol{\theta}']) = \mathrm{dist}^2(\boldsymbol{\theta}_\mathrm{c}, \boldsymbol{\theta}_\mathrm{c}') = \sum_{i=1}^N \|\theta_{\mathrm{c},i} - \theta_{\mathrm{c},i}'\|^2. \tag{60}$$

It follows that $\mathrm{dist}^2([\boldsymbol{\theta}], [\boldsymbol{\theta}']) = \mathrm{dist}^2(\boldsymbol{\theta}_\mathrm{c}, \boldsymbol{\theta}_\mathrm{c}')$, hence the mapping $[\boldsymbol{\theta}] \mapsto \boldsymbol{\theta}_\mathrm{c}$ is an isometry between $\bar{\mathcal{P}}$ and the Euclidean space $\mathcal{P}$. We thus conclude that $\bar{\mathcal{P}}$ is a flat manifold and that its curvature tensor vanishes identically [15, Chap. 7]. Theorem 3 and Proposition 6 then yield:

$$\mathbb{E}\left\{(\hat{\boldsymbol{\theta}}_\mathrm{c} - \boldsymbol{\theta}_\mathrm{c})(\hat{\boldsymbol{\theta}}_\mathrm{c} - \boldsymbol{\theta}_\mathrm{c})^\top\right\} \triangleq C \succeq \mathcal{L}^\dagger \otimes \Sigma, \text{ and} \tag{61}$$

$$\mathbb{E}\Big\{\sum_{i=1}^N \|\hat{\theta}_{\mathrm{c},i} - \theta_{\mathrm{c},i}\|^2\Big\} \geq \mathrm{trace}(\mathcal{L}^\dagger)\mathrm{trace}(\Sigma). \tag{62}$$

We now interpret the CRB (61). 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 (61) 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 $\hat{\boldsymbol{\theta}}$ nor $\boldsymbol{\theta}$ 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:

$$\mathbb{E}\left\{\left\|(\hat{\theta}_i - \hat{\theta}_j) - (\theta_i - \theta_j)\right\|^2\right\} \geq (e_i - e_j)^\top \mathcal{L}^\dagger (e_i - e_j) \cdot \mathrm{trace}(\Sigma). \tag{63}$$

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 [20]. 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 [20] 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. 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 $\max L(\hat{\boldsymbol{\theta}})$ (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.

B. Synchronization of rotations

We now consider a second example, closely related to the first one. Let $\mathrm{SO}(3) = \{R \in \mathbb{R}^{3\times 3} : R^\top R = I_3, \det(R) = 1\}$ denote the set of rotation matrices in $\mathbb{R}^3$. Let $\mathbf{R} = (R_1, \ldots, R_N)$ denote a set of unknown but deterministic rotations: those are the parameters. The natural parameter space is thus $\mathcal{P} = \mathrm{SO}(3)^N$. Consider an undirected graph on N nodes with edge set $\mathcal{E}$ as in the previous example. For each edge $\{i, j\} \in \mathcal{E}$, $i < j$, we are given a measurement of the relative rotation between R_i and R_j:

$$H_{ij} = R_i R_j^\top + \sigma N_{ij}, \tag{64}$$

where $\sigma > 0$ and N_{ij} is a 3×3 matrix whose entries are i.i.d. sampled from $\mathcal{N}(0, 1)$. Let us also assume that the N_{ij}'s are independent from each other. The proposed noise model could be disputed. Indeed, measurements of relative rotations should be rotation matrices too. In contrast, in the proposed model, the H_{ij}'s are arbitrary 3×3 matrices. We conduct the calculations with this noise model here for it admits a concise exposition. A more compelling model will be considered in [17].

The task is to estimate the rotations $(R_i)_{i=1,\ldots,N}$ from the measurements $(H_{ij})_{\{i<j\}\in\mathcal{E}}$. We expect a singular FIM since the measurements only convey information about relative rotations, not absolute rotations. Indeed, all sets of rotations of the form $\mathbf{R}Q \overset{\text{not.}}{=} (R_1 Q, \ldots, R_N Q)$ for arbitrary $Q \in \mathrm{SO}(3)$ lead to the same probability distribution function for the measurements.

We will need a few tools to work on $\mathrm{SO}(3)$, see [21]. The tangent space at $R \in \mathrm{SO}(3)$ is the set $\mathrm{T}_R\mathrm{SO}(3) = \{R\Omega : \Omega \in \mathfrak{so}(3)\}$, where $\mathfrak{so}(3) = \{\Omega \in \mathbb{R}^{3\times 3} : \Omega^\top = -\Omega\}$ is the set of skew-symmetric matrices. It is endowed with the usual metric $\langle R\Omega, R\Omega' \rangle_R = \mathrm{trace}(\Omega^\top \Omega')$, turning it into a Riemannian submanifold of $\mathbb{R}^{3\times 3}$. The logarithmic map between two rotation matrices R and R' is given by the matrix logarithm $\log : \mathrm{SO}(3) \to \mathfrak{so}(3)$ as $\mathrm{Log}_R(R') = R \log(R^\top R')$. The Riemannian distance follows as

$$\mathrm{dist}(R, R') = \left\| \log(R^\top R') \right\|_\mathrm{F} = \sqrt{2}\alpha, \tag{65}$$

where $\alpha \in [0, \pi]$ is the angle by which the aligning rotation $R^\top R'$ rotates. By element-wise extension, the tangent space to $\mathcal{P}$ at $\mathbf{R}$ is $\mathrm{T}_\mathbf{R}\mathcal{P} = \{\mathbf{R\Omega} \overset{\text{not.}}{=} (R_1\Omega_1, \ldots, R_N\Omega_N) : \mathbf{\Omega} = (\Omega_1, \ldots, \Omega_N) \in \mathfrak{so}(3)^N\}$. The inner product is $\langle \mathbf{R\Omega}, \mathbf{R\Omega}' \rangle_\mathbf{R} = \sum_{i=1}^N \mathrm{trace}(\Omega_i^\top \Omega_i')$.

The probability distribution function of the random matrix $\sigma N_{ij} = H_{ij} - R_i R_j^\top$ is proportional to $\exp(-\|N_{ij}\|_\mathrm{F}^2 / (2\sigma^2))$, where $\|\cdot\|_\mathrm{F}$ denotes the Frobenius norm: $\|A\|_\mathrm{F}^2 = \langle A, A \rangle_\mathrm{F} = \mathrm{trace}(A^\top A)$. Hence, up to unimportant additive constants, the log-likelihood function at $\hat{\mathbf{R}} = (\hat{R}_1, \ldots, \hat{R}_N) \in \mathcal{P}$ is:

$$L(\hat{\mathbf{R}}) = -\frac{1}{2\sigma^2} \sum_{\{i<j\}\in\mathcal{E}} \left\| H_{ij} - \hat{R}_i \hat{R}_j^\top \right\|_\mathrm{F}^2. \tag{66}$$

It is easy to compute the directional derivative of the log-likelihood at $\mathbf{R}$ along the tangent vector $\mathbf{R}\boldsymbol{\Omega}$:

$$\mathrm{D}L(\mathbf{R})[\mathbf{R}\boldsymbol{\Omega}] = \frac{1}{\sigma} \sum_{\{i<j\}\in\mathcal{E}} \left\langle N_{ij}, R_i(\Omega_i - \Omega_j)R_j^\top \right\rangle_\mathrm{F}. \tag{67}$$

Hence, using independence of the entries of the N_{ij}'s and the fact that the N_{ij}'s have zero-mean, the Fisher information metric at $\mathbf{R}$ is given by:

$$\begin{aligned} \mathbf{F}(\mathbf{R})[\mathbf{R}\boldsymbol{\Omega}, \mathbf{R}\boldsymbol{\Omega}] &= \mathbb{E}\left\{ (\mathrm{D}L(\mathbf{R})[\mathbf{R}\boldsymbol{\Omega}])^2 \right\} \\ &= \frac{1}{\sigma^2} \sum_{\{i<j\}\in\mathcal{E}} \mathbb{E}\left\{ \left\langle N_{ij}, R_i(\Omega_i - \Omega_j)R_j^\top \right\rangle_\mathrm{F}^2 \right\} \\ &= \frac{1}{\sigma^2} \sum_{\{i<j\}\in\mathcal{E}} \left\| R_i(\Omega_i - \Omega_j)R_j^\top \right\|_\mathrm{F}^2 \\ &= \frac{1}{\sigma^2} \sum_{\{i<j\}\in\mathcal{E}} \|\Omega_i - \Omega_j\|_\mathrm{F}^2. \end{aligned} \tag{68}$$

The FIM is thus independent of the true rotations $\mathbf{R}$. From now on, we will write $\mathbf{F}$ instead of $\mathbf{F}(\mathbf{R})$. To represent $\mathbf{F}$ as a matrix, we need to pick an orthonormal basis of $\mathrm{T}_\mathbf{R}\mathcal{P}$. Let E_1, E_2, E_3 form an orthonormal basis of $\mathfrak{so}(n)$. This choice leads to an orthonormal basis for the tangent space $\mathrm{T}_\mathbf{R}\mathcal{P}$:

$$(e_{ik})_{i=1,\ldots,N,k=1,\ldots,3}, \text{ with } e_{ik} = (0, \ldots, 0, R_iE_k, 0, \ldots, 0),$$

a zero vector except for the i^th component equal to R_iE_k.

The matrix F which represents the quadratic form $\mathbf{F}$ w.r.t. this basis is a $3N \times 3N$ symmetric matrix composed of 3×3 blocks. $F_{ij,k\ell}$ refers to the (k, ℓ) entry in the (i, j) block. The value of this entry,

$$F_{ij,k\ell} = \mathbf{F}[e_{ik}, e_{j\ell}], \tag{69}$$

is best obtained via the polarization identity (42) applied to equation (68). For example, let us consider $\{i < j\} \in \mathcal{E}$ and $k, \ell \in \{1, 2, 3\}$. Polarization yields:

$$F_{ij,k\ell} = \frac{1}{4\sigma^2} \left(\|E_k - E_\ell\|_\mathrm{F}^2 - \|E_k + E_\ell\|_\mathrm{F}^2 \right). \tag{70}$$

For $k = \ell$, this evaluates to $-1/\sigma^2$. For $k \neq \ell$, this evaluates to zero. Hence, the (i, j) block of F is $(-1/\sigma^2)I_3$ if i and j are adjacent nodes. Similar algebra establishes that if nodes i and j are not adjacent, the corresponding block is zero. Similarly, the i^{th} diagonal block of F has the form $(|V_i|/\sigma^2)I_3$, where $|V_i|$ is the degree of node i. We thus obtain the same Laplacian structure as in the first example:

$$F = \frac{1}{\sigma^2}(\mathcal{L} \otimes I_3). \tag{71}$$

F is indeed rank-deficient. As in the previous example, for connected graphs, we have $\ker F = \{\mathbb{1}_N \otimes t : t \in \mathbb{R}^3\}$. These vectors correspond to the rotation of all R_i's in the same direction, which indeed leaves relative rotations unaffected.

c) With anchors: Let us consider $A \subset \{1, \ldots, N\}$, $A \neq \emptyset$, such that all R_i with $i \in A$ are known; these are anchors. The resulting parameter space $\bar{\mathcal{P}} = \{\hat{\mathbf{R}} \in \mathcal{P} : \hat{\mathrm{R}}_i = \mathrm{R}_i \ \forall i \in A\}$ is a Riemannian submanifold of $\mathcal{P}$.

Let $\hat{\mathbf{R}}$ be an unbiased estimator for the anchored problem. As defined by equation (5), the covariance matrix of that estimator w.r.t. the basis of e_{ik}'s is a $3N \times 3N$ matrix with blocks of size 3×3. Indexing entry (k, ℓ) in the block (i, j) as $C_{ij,k\ell}$, we have by definition:

$$C_{ij,k\ell} = \mathbb{E}\left\{\langle X, e_{ik}\rangle_{\mathbf{R}} \langle X, e_{j\ell}\rangle_{\mathbf{R}}\right\}, \tag{72}$$

where $X = \mathrm{Log}_{\mathbf{R}}(\hat{\mathbf{R}}) = \Big(\mathrm{Log}_{R_1}(\hat{R}_1), \ldots, \mathrm{Log}_{R_N}(\hat{R}_N)\Big)$ is the (random) error vector. In particular, the trace of C is the expected squared norm of X, that is, the expected squared distance between the true rotations $\mathbf{R}$ and the estimator $\hat{\mathbf{R}}$.

The orthogonal projector from $\mathrm{T}_{\mathbf{R}}\mathcal{P}$ to $\mathrm{T}_{\mathbf{R}}\bar{\mathcal{P}}$ sets all components of a tangent vector corresponding to anchored nodes to zero. Hence, as in the previous example, we define $\mathcal{L}_A$ to be the Laplacian with rows and columns corresponding to anchored nodes forced to zero. Theorem 2 yields the anchored CRB:

$$C \succeq \sigma^2 \mathcal{L}_A^{\dagger} \otimes I_3 + \text{curvature terms}. \tag{73}$$

This bound lends itself to a similar interpretation as in the previous example.

At large SNR, we expect the curvature terms to become negligible. Unfortunately, invoking the simple argument (32) is not sufficient. Indeed, SO(3) has constant sectional curvature of 1/8. Hence, a (tight) upperbound on the absolute value of the sectional curvatures of $\bar{\mathcal{P}}$ is $K_{\max} = 1/8$. Requiring $\mathrm{dist}(\mathbf{R}, \hat{\mathbf{R}}) \ll \sqrt{8}$ is too restrictive when a large number N of rotations have to be estimated. In [17, Appendix C], the curvature terms for $\bar{\mathcal{P}}$ are derived and it is shown that

$$R_m(\sigma^2 \mathcal{L}_A^{\dagger} \otimes I_3) = \frac{3\sigma^2}{4} \mathrm{ddiag}(\mathcal{L}_A^{\dagger}) \otimes I_3, \tag{74}$$

where ddiag puts all off-diagonal entries of a matrix to zero. Hence, Theorem 4 yields the following CRB:

$$C \succeq \sigma^2 \mathcal{L}_A^\dagger \otimes I_3 + \frac{3\sigma^4}{4}\left(\mathrm{ddiag}(\mathcal{L}_A^\dagger)\mathcal{L}_A^\dagger + \mathcal{L}_A^\dagger \mathrm{ddiag}(\mathcal{L}_A^\dagger)\right) \otimes I_3. \tag{75}$$

As expected, for large SNR, that is, for small σ^2, the curvature terms are negligible.

Taking traces on both sides of (73), we obtain an inequality for the total variance (neglecting curvature):

$$\mathbb{E}\left\{\mathrm{dist}^2(\hat{\mathbf{R}}, \mathbf{R})\right\} \triangleq \mathbb{E}\Big\{\sum_{i=1}^N \mathrm{dist}^2(\hat{R}_i, R_i)\Big\} \geq 3\sigma^2 \mathrm{trace}(\mathcal{L}_A^\dagger).$$

d) Without anchors: If there are no anchors, we work on the quotient space $\bar{\mathcal{P}} = \mathcal{P}/\sim$, where $\mathbf{R} \sim \mathbf{R}'$ iff there exists a rotation $Q \in \mathrm{SO}(3)$ such that $\mathbf{R}' = \mathbf{R}Q \overset{\text{not.}}{=} (R_1 Q, \ldots, R_N Q)$. The distance between the equivalence classes $[\mathbf{R}]$ and $[\mathbf{R}']$ on $\bar{\mathcal{P}}$ is the distance between their best aligned members:

$$\mathrm{dist}^2([\mathbf{R}], [\mathbf{R}']) = \min_{Q \in \mathrm{SO}(3)} \mathrm{dist}^2(\mathbf{R}Q, \mathbf{R}') = \min_{Q \in \mathrm{SO}(3)} \sum_{i=1}^N \left\| \log(Q^\top R_i^\top R_i') \right\|_\mathrm{F}^2. \tag{76}$$

This is well defined since $\mathrm{SO}(3)$ is compact. Computing the optimal Q amounts to computing a Karcher mean on $\mathrm{SO}(3)$, a problem for which good algorithms are available [17], [22], [23].

A simple application of Theorem 3 then yields the following bound on the variance of an unbiased estimator $\hat{\mathbf{R}}$ for the anchor-free synchronization problem:

$$\mathbb{E}\left\{\mathrm{dist}^2([\mathbf{R}], [\hat{\mathbf{R}}])\right\} \geq 3\sigma^2 \mathrm{trace}(\mathcal{L}^\dagger) + \text{curvature terms}.$$

A more complete study of the anchor-free case is conducted in [17]. Among other things, it is shown there that the curvature terms for the quotient space $\bar{\mathcal{P}}$ are on the order of $\mathcal{O}(\sigma^4(1 + 1/N))$. Hence, they remain negligible for large SNR, as expected.

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 two simple examples how these theorems easily provide meaningful bounds for estimation problems with indeterminacies, whether these are dealt with by including prior knowledge in the form of constraints or by acknowledging the quotient nature of the parameter space. We also observed on these same examples that fixing indeterminacies by adding constraints results in different CRB's than if the quotient nature is acknowledged.

ACKNOWLEDGMENTS

NB gratefully acknowledges J. Xavier and V. Barroso for figures 2–3 as well as P.-A. Absil, A. Singer and the anonymous referees 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 A
FUNDAMENTALS OF DIFFERENTIAL GEOMETRY

In this appendix, we give an informal review of the elementary tools of differential geometry used in the present paper. A number of standard textbooks cover all the necessary concepts [13], [16], [24]. The monograph [14] features an algorithmically-oriented introduction to differential geometry and is available for free on the publisher's website. The footnotes in [3] can also constitute a refresher.

A *smooth manifold* is a set $\mathcal{P}$ that locally "looks like" $\mathbb{R}^d$, that is, $\mathcal{P}$ can locally be charted by a choice of smooth coordinates $(x_1, \ldots, x_d)$, $x_i : \Omega \subset \mathcal{P} \to \mathbb{R}$, up to some natural consistency conditions for charts with overlapping domains. A trivial example is $\mathbb{R}^d$ itself. A simple example is the sphere $\mathbb{S}^2 = \{\theta \in \mathbb{R}^3 : \theta^\top\theta = 1\}$. Locally, $\mathbb{S}^2$ can be mapped to an open set of $\mathbb{R}^2$ in a smooth way. The *dimension* of the manifold is the number of coordinates required: $\dim \mathcal{P} = d$. A curve on $\mathcal{P}$, $c : \mathbb{R} \to \mathcal{P}$, is smooth if it is smooth when expressed in the charts.

There exist many different manifolds of practical interest for signal processing applications. We cite just a few. The (compact) Stiefel manifold $\mathrm{St}(n, p)$ is the set of $n \times p$ orthonormal matrices [14]. Special cases include the sphere $\mathbb{S}^{n-1}$ ($p = 1$) and the orthogonal group $\mathrm{O}(n)$ ($p = n$). The Grassmann manifold $\mathrm{Gr}(n, p)$ is the set of vector subspaces of dimension p embedded in $\mathbb{R}^n$ [14]. The special orthogonal group $\mathrm{SO}(n)$ is the set of $n \times n$ orthogonal matrices with determinant 1, which corresponds to the set of rotations in $\mathbb{R}^n$ [21].

At each point $\theta \in \mathcal{P}$, we define a *tangent space* $\mathrm{T}_\theta\mathcal{P}$, which is a vector space of dimension d. For embedded manifolds such as $\mathbb{S}^2$, the tangent spaces correspond to the usual concept of tangent planes to $\mathbb{S}^2$ in $\mathbb{R}^3$: $\mathrm{T}_\theta\mathbb{S}^2 = \{u \in \mathbb{R}^3 : \theta^\top u = 0\}$. In a more general setting, it is desirable to define $\mathrm{T}_\theta\mathcal{P}$ without referencing an embedding space. This is done by considering smooth curves on $\mathcal{P}$ that pass through θ and grouping the curves that do so with the same "velocity" (derivative as expressed in the charts) into an equivalence class, which is then called a tangent vector.

Let $L : \mathcal{P} \to \mathbb{R}$ be a smooth scalar field over $\mathcal{P}$. The *directional derivative* of L at θ along a tangent vector u is defined by $\mathrm{D}L(\theta)[u] = (L \circ c)'(0)$, the derivative of $L \circ c : \mathbb{R} \to \mathbb{R}$ at $t = 0$, where $\circ$ denotes function composition and $c : \mathbb{R} \to \mathcal{P}$ is a smooth curve passing through θ at $t = 0$ with velocity u. For the sphere for example, this correspondence amounts to $c'(0) = u$, where c is seen as a function from $\mathbb{R}$ to $\mathbb{R}^3$. If L is smoothly defined in an open neighborhood of $\mathbb{S}^2$ in $\mathbb{R}^3$, this reduces to a classical derivative: $\mathrm{D}L(\theta)[u] = \lim_{t\to 0}(L(\theta + tu) - L(\theta))/t$.

Each tangent space may be endowed with a metric (an inner product) $\langle\cdot,\cdot\rangle_\theta : \mathrm{T}_\theta\mathcal{P} \times \mathrm{T}_\theta\mathcal{P} \to \mathbb{R}$. The associated norm is defined by $\|u\|_\theta^2 = \langle u, u\rangle_\theta$. When this metric varies continuously as a function of θ, $\mathcal{P}$ is a *Riemannian manifold*. The gradient of a scalar field L at θ, noted $\mathrm{grad}\, L(\theta)$, gives the steepest-ascent direction of L at θ. By definition, it is the unique tangent vector that satisfies the following:

$$\forall u \in \mathrm{T}_\theta\mathcal{P}, \quad \langle \mathrm{grad}\, L(\theta), u\rangle_\theta = \mathrm{D}L(\theta)[u]. \tag{77}$$

Notice that, while the notion of directional derivative does not depend on a particular choice of metric, the notion of gradient does.

A curve $\gamma : I \subset \mathbb{R} \to \mathcal{P}$ is a *geodesic* if it has zero acceleration. A proper definition of acceleration along a curve requires the concept of *affine connection* ∇, which we omit here. For each $\theta \in \mathcal{P}$ and $u \in \mathrm{T}_\theta\mathcal{P}$, there exists a unique geodesic γ_u such that $\gamma_u(0) = \theta$ and $\gamma_u'(0) = u$. The *exponential map* at θ is the mapping $\mathrm{Exp}_\theta : \mathrm{T}_\theta\mathcal{P} \to \mathcal{P}$ defined by $\mathrm{Exp}_\theta(u) = \gamma_u(1)$. The *logarithmic map* is the principal inverse of the exponential map, $\mathrm{Log}_\theta : \mathcal{P} \to \mathrm{T}_\theta\mathcal{P}$, that is, $u = \mathrm{Log}_\theta(\theta')$ is the shortest tangent vector such that $\mathrm{Exp}_\theta(u) = \theta'$. The notions of geodesic, exponential map and logarithmic map are usually only defined locally. For $\mathcal{P} = \mathbb{R}^d$ with the usual metric $\langle u, v\rangle_\theta = \mathrm{trace}(u^\top v)$, geodesics are straight lines, $\mathrm{Exp}_\theta(u) = \theta + u$ and $\mathrm{Log}_\theta(\theta') = \theta' - \theta$. On the sphere $\mathbb{S}^2$, geodesics are great circles, such as the equator. The *Riemannian distance* or *geodesic distance* between two points $\theta, \theta' \in \mathcal{P}$, is the length of the shortest geodesic arc connecting the two points: $\mathrm{dist}(\theta, \theta') = \|\mathrm{Log}_\theta(\theta')\|_\theta$. On the unit sphere, the Riemannian distance is the arc length, or angle in radians, separating two points.

A *flat manifold* is a manifold that is locally isometric to a Euclidean space $\mathbb{R}^d$. For example, any one-dimensional smooth manifold is flat. Think of the circle $\mathbb{S}^1 = \{\theta \in \mathbb{R}^2 : \theta^\top\theta = 1\}$ as a Riemannian submanifold of $\mathbb{R}^2$ with its usual metric: it is always possible to cut out a bit of $\mathbb{S}^1$ and straighten it without distorting distances. Similarly, the torus $\mathbb{S}^1 \times \mathbb{S}^1$ is flat, being the product of two circles. This example may contrast with our intuition if we think of the torus as an embedded manifold of $\mathbb{R}^3$: their topologies are the same, but their Riemannian structures are not. On the other hand, the sphere $\mathbb{S}^2$ is not flat: attempting to flatten an orange peel will necessarily result in tearing, stretching or compression.

Given two tangent vectors $u, v \in \mathrm{T}_\theta\mathcal{P}$ (not co-linear), the *sectional curvature* $K(u, v)$ is a real number that measures the curvature of $\mathcal{P}$ along the 2-dimensional subspace of $\mathrm{T}_\theta\mathcal{P}$ spanned by u and v. For example, on $\mathbb{S}^2$, all sectional curvatures are equal to 1. The *Riemannian curvature tensor* R is another (but equivalent) means of quantifying the "non-flatness" of a manifold. It is related to the sectional curvatures via $K(u, v) = \langle R(u, v)v, u\rangle_\theta / (\|u\|_\theta^2\|v\|_\theta^2 - \langle u, v\rangle_\theta^2)$. For flat manifolds, curvature is zero.

Let $\mathcal{P}$ be a Riemannian manifold and let $\bar{\mathcal{P}}$ be a submanifold of $\mathcal{P}$, i.e., $\bar{\mathcal{P}}$ is embedded in $\mathcal{P}$, like the sphere is embedded in 3-space. The tangent space $\mathrm{T}_\theta\bar{\mathcal{P}}$ is a vector subspace of $\mathrm{T}_\theta\mathcal{P}$. $\bar{\mathcal{P}}$ is turned into a *Riemannian submanifold* of $\mathcal{P}$ by giving it a Riemannian metric that is the restriction of the Riemannian metric of $\mathcal{P}$. For example, the classical metric on $\mathbb{R}^3$ is $\langle u, v\rangle = u^\top v$. By defining $\langle u, v\rangle_\theta = u^\top v$ on $\mathrm{T}_\theta\mathbb{S}^2$, we equip $\mathbb{S}^2$ with a Riemannian submanifold structure. The orthogonal projector from $\mathrm{T}_\theta\mathcal{P}$ to $\mathrm{T}_\theta\bar{\mathcal{P}}$ is noted P_θ. For the sphere, it is given by $\mathrm{P}_\theta u = (I - \theta\theta^\top)u$.

Let $\mathcal{P}$ be a Riemannian manifold and let $\sim$ be an equivalence relation on $\mathcal{P}$. Let $[\theta] = \{\theta' \in \mathcal{P} : \theta' \sim \theta\}$ be the equivalence class of θ and let $\bar{\mathcal{P}} = \{[\theta] : \theta \in \mathcal{P}\}$ be the set of equivalence classes. Let us further assume that $\bar{\mathcal{P}}$ is a smooth manifold, i.e., it can be smoothly charted. Then, the projection $\pi : \mathcal{P} \to \bar{\mathcal{P}}$ is a *submersion* if its differential at θ, $\mathrm{D}\pi(\theta) : \mathrm{T}_\theta\mathcal{P} \to \mathrm{T}_{[\theta]}\bar{\mathcal{P}}$, is a surjective map for all θ. In that case, $\bar{\mathcal{P}}$ is a *quotient manifold* $\mathcal{P}/\sim$, because we "quotient out" the equivalence relation from $\mathcal{P}$. The equivalence classes $\pi(\theta)$ are termed the *fibers*; they are Riemannian submanifolds of $\mathcal{P}$. As such, they admit tangent spaces, called the *vertical spaces*: $\mathrm{V}_\theta = \mathrm{T}_\theta\pi(\theta) = \ker \mathrm{D}\pi(\theta)$. The vertical space is a vector subspace of $\mathrm{T}_\theta\mathcal{P}$. The orthogonal complement to V_θ in $\mathrm{T}_\theta\mathcal{P}$ is called the *horizontal space*, $\mathrm{H}_\theta = \{u \in \mathrm{T}_\theta\mathcal{P} : \langle u, v\rangle_\theta = 0 \,\forall v \in \mathrm{V}_\theta\}$. If furthermore $\bar{\mathcal{P}}$ is endowed with a Riemannian metric and $\mathrm{D}\pi(\theta)$ restricted to H_θ is an isometry, in the sense that

$$\forall\theta \in \mathcal{P}, \forall u, v \in \mathrm{H}_\theta, \quad \langle \mathrm{D}\pi(\theta)[u], \mathrm{D}\pi(\theta)[v]\rangle_{[\theta]} = \langle u, v\rangle_\theta\,,$$

then π is a *Riemannian submersion* and $\bar{\mathcal{P}}$ is a *Riemannian quotient manifold* of $\mathcal{P}$. Being an isometry, $\mathrm{D}\pi(\theta)$ establishes a natural correspondence between the abstract tangent vectors in $\mathrm{T}_{[\theta]}\bar{\mathcal{P}}$ and the often more concrete horizontal vectors. The *horizontal lift* of a tangent vector $X \in \mathrm{T}_{[\theta]}\bar{\mathcal{P}}$ to H_θ is the unique horizontal vector $\xi \in \mathrm{H}_\theta$ that verifies $\mathrm{D}\pi(\theta)[\xi] = X$. The definition of Riemannian submersion ensures that this is well-defined. The orthogonal projector from the total tangent space $\mathrm{T}_\theta\mathcal{P}$ to H_θ is here noted $P\mathrm{H}_\theta$.

APPENDIX B

PSEUDO-INVERSE AND KRONECKER PRODUCT COMMUTE

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

$$A^\dagger \otimes B^\dagger = (V_A \Sigma_A^{-1} U_A^\top) \otimes (V_B \Sigma_B^{-1} U_B^\top) \tag{78}$$

$$= (V_A \otimes V_B)(\Sigma_A^{-1} \otimes \Sigma_B^{-1})(U_A \otimes U_B)^\top, \tag{79}$$

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

$$A \otimes B = (U_A \otimes U_B)(\Sigma_A \otimes \Sigma_B)(V_A \otimes V_B)^\top. \tag{80}$$

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

$$(A \otimes B)^\dagger = (V_A \otimes V_B)(\Sigma_A \otimes \Sigma_B)^{-1}(U_A \otimes U_B)^\top. \tag{81}$$

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] J. Gorman and A. Hero, “Lower bounds for parametric estimation with constraints,” *Information Theory, IEEE Transactions on*, vol. 36, no. 6, pp. 1285–1301, nov 1990.

[7] Z. Ben-Haim and Y. Eldar, “On the constrained Cramér-Rao bound with a singular Fisher information matrix,” *Signal Processing Letters, IEEE*, vol. 16, no. 6, pp. 453 –456, june 2009.

[8] 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.

[9] P. Stoica and T. Marzetta, “Parameter estimation problems with singular information matrices,” *Signal Processing, IEEE Transactions on*, vol. 49, no. 1, pp. 87–90, 2001.

[10] 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.

[11] L. Scharf and L. McWhorter, “Geometry of the Cramer-Rao bound,” *Signal Processing*, vol. 31, no. 3, pp. 301 – 311, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/016516849390088R

[12] S. Amari and H. Nagaoka, *Methods of information geometry*. American Mathematical Society, 2007, vol. 191.

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

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

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

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

[17] 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. 2012.08, October 2012.

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

[19] 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.

[20] 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.

[21] D. Bump, *Lie groups*, ser. Graduate Texts in Mathematics. Springer, 2004, vol. 225.

[22] Q. Rentmeesters and P.-A. Absil, “Algorithm comparison for Karcher mean computation of rotation matrices and diffusion tensors,” in *EUSIPCO 2011: 19th European Signal Processing Conference*, 2011, pp. 2229–2233.

[23] J. Manton, “A globally convergent numerical algorithm for computing the centre of mass on compact lie groups,” in *Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th*, vol. 3. IEEE, 2004, pp. 2211–2216.

[24] W. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, ser. Pure and Applied Mathematics. Elsevier, 1986, vol. 120.