Covariance, Subspace, and Intrinsic Cramér–Rao Bounds

Steven Thomas Smith, *Senior Member, IEEE*

***Abstract*—Cramér–Rao bounds on estimation accuracy are established for estimation problems on arbitrary manifolds in which no set of intrinsic coordinates exists. The frequently encountered examples of estimating either an unknown subspace or a covariance matrix are examined in detail. The set of subspaces, called the Grassmann manifold, and the set of covariance (positive-definite Hermitian) matrices have no fixed coordinate system associated with them and do not possess a vector space structure, both of which are required for deriving classical Cramér–Rao bounds. Intrinsic versions of the Cramér–Rao bound on manifolds utilizing an arbitrary affine connection with arbitrary geodesics are derived for both biased and unbiased estimators. In the example of covariance matrix estimation, closed-form expressions for both the intrinsic and flat bounds are derived and compared with the root-mean-square error (RMSE) of the sample covariance matrix (SCM) estimator for varying sample support K. The accuracy bound on unbiased covariance matrix estimators is shown to be about $(10/\log 10)n/K^{1/2}$ dB, where n is the matrix order. Remarkably, it is shown that from an intrinsic perspective, *the SCM is a biased and inefficient estimator* and that the bias term reveals the dependency of estimation accuracy on sample support observed in theory and practice. The RMSE of the standard method of estimating subspaces using the singular value decomposition (SVD) is compared with the intrinsic subspace Cramér–Rao bound derived in closed form by varying both the signal-to-noise ratio (SNR) of the unknown p-dimensional subspace and the sample support. In the simplest case, the Cramér–Rao bound on subspace estimation accuracy is shown to be about $(p(n-p))^{1/2}K^{-1/2}\mathrm{SNR}^{-1/2}$ rad for p-dimensional subspaces. It is seen that the SVD-based method yields accuracies very close to the Cramér–Rao bound, establishing that the principal invariant subspace of a random sample provides an excellent estimator of an unknown subspace. The analysis approach developed is directly applicable to many other estimation problems on manifolds encountered in signal processing and elsewhere, such as estimating rotation matrices in computer vision and estimating subspace basis vectors in blind source separation.**

***Index Terms*—Adaptive arrays, adaptive estimation, adaptive signal processing, covariance matrices, differential geometry, error analysis, estimation, estimation bias, estimation efficiency, Fisher information, Grassmann manifold, homogeneous space, matrix decomposition, maximum likelihood estimaiton, natural gradient, nonlinear estimation, parameter estimation, parameter space methods, positive definitive matrices, Riemannian curvature, Riemannian manifold, singular value decomposition.**

I. Introduction

Estimation problems are typically posed and analyzed for a set of fixed parameters, such as angle and Doppler. In contrast, estimation problems on manifolds, where no such set of intrinsic coordinates exists, are frequently encountered in signal processing applications and elsewhere. Two common examples are estimating either an unknown covariance matrix or a subspace. Because neither the set of covariance (positive definite Hermitian) matrices nor the set of subspaces (the Grassmann manifold) are equipped with an intrinsic coordinate system or a vector space structure, classical Cramér–Rao bound (CRB) analysis [17], [55] is not directly applicable. To address this class of problems, an intrinsic treatment of Cramér–Rao analysis specific to signal processing problems is established here. Intrinsic versions of the CRB have also been developed for Riemannian manifolds [34], [43], [44], [46], [51], [52], [68], [79], [80], statistical manifolds [5], and for the application of quantum inference [11], [53].

The original contributions of this paper are 1) a derivation of biased and unbiased intrinsic CRBs for signal processing and related fields; 2) a new proof of the CRB (Theorem 2) that connects the inverse of the Fisher information matrix with its appearance in the (natural) gradient of the log-likelihood function; 3) several results [Theorem 4, Corollary 5, (143)] that bound the estimation accuracy of an unknown covariance matrix or subspace; 4) the noteworthy discovery that from an intrinsic perspective, the sample covariance matrix (SCM) is a biased and inefficient estimator (Theorem 7), and the fact that the bias corresponds to the SCMs poor estimation quality at low sample support (Corollary 5)—this contradicts the well-known fact that $E[\hat{\mathbf{R}}] = \mathbf{R}$ because the linear expectation operator implicitly treats the covariance matrices as a convex cone included in the vector space $\mathbb{R}^{n^2}$, compared to the intrinsic treatment of the covariance matrices in this paper; 5) a generalization of the expression for Fisher information (Theorem 1) that employs the Hessian of the log-likelihood function for arbitrary affine connections—a useful tool because in the great majority of applications the second-order terms are much easier to compute; 6) a geometric treatment of covariance matrices as the quotient space $P_n \cong \boldsymbol{Gl}(n,\mathbb{C})/\boldsymbol{U}(n)$ (i.e., the Hermitian part of the matrix polar decomposition), including a natural distance between covariance matrices that has not appeared previously in the signal processing literature; and 7) a comparison between the accuracy of the standard subspace estimation method employing the singular value decomposition (SVD) and the CRB for subspace estimation.

In contrast to previous literature on intrinsic Cramér–Rao analysis, it is shown explicitly how to compute practical estimation bounds on a parameter space defined by a manifold, independently of any particular metric or affine structure. As elsewhere, the standard approach is used to generalize classical bounds to Riemannian manifolds via the exponential map, i.e., geodesics emanating from the estimate to points in the parameter space. Just as with classical bounds, the unbiased intrinsic

Manuscript received June 23, 2003; revised June 3, 2004. This work was supported by the United States Air Force under Air Force Contract F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Randolph Moses.

The author is with the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington MA 02420 USA (e-mail: stsmith@ll.mit.edu).

Digital Object Identifier 10.1109/TSP.2005.845428

CRBs depend asymptotically only on the Fisher information and do not depend in any nontrivial way on the choice of measurement units, e.g., feet versus meters. Though the mathematical concepts used throughout the paper are well known to the differential geometry community, a brief informal background of the key ideas is provided in footnotes for readers unfamiliar with some of the technicalities, as is a table at the end of the paper, comparing the general concepts to their more familiar counterparts in $\mathbb{R}^n$.

The results developed in this paper are general enough to be applied to the numerous estimation problems on manifolds that appear in the literature. Zheng and Tse [82], in a geometric approach to the analysis of Hochwald and Marzetta [35], compute channel capacities for communications problems with an unknown propagation gain matrix represented by an element on the Grassmann manifold. Grenander *et al.* [30] derive Hilbert–Schmidt lower bounds for estimating points on a Lie group for automatic target recognition. Srivastava [70] applies Bayesian estimation theory to subspace estimation, and Srivastava and Klassen [71], [72] develop an extrinsic approach to the problem of estimating points on a manifold, specifically Lie groups and their quotient spaces (including the Grassmann manifold), and apply their method to estimating target pose. Bhattacharya and Patrangenaru [9] treat the general problem of estimation on Riemannian manifolds. Estimating points on the rotation group, or the finite product space of the rotation group, occurs in several diverse applications. Ma *et al.* [41] describe a solution to the motion recovery problem in computer vision, and Adler *et al.* [2] use a set of rotations to describe a model for the human spine. Douglas *et al.* [20], Smith [66], and many others [21]–[23], develop gradient-based adaptive algorithms using the natural metric structure of a constrained space. For global estimation bounds rather than the local ones developed in this paper, Rendas and Moura [59] define a general ambiguity function for parameters in a statistical manifold. In the area of blind signal processing, Cichocki *et al.* [14], [15], Douglas [19], and Rahbar and Reilly [54] solve estimation problems on the Stiefel manifold to accomplish blind source separation, and Xavier [79] analyzes blind MIMO system identification problems using an intrinsic approach. Lo and Willsky [39] analyze optimal estimation problems on Abelian Lie groups. Readers may also be interested in the central role played by geometrical statistical analysis in establishing Wegener's theory of continental drift [18], [27], [43].

A. Model Estimation Problem

Consider the problem [68] of estimating the unknown n by p matrix $\mathbf{Y}$, $n \geq p$, given the statistical model

$$\mathbf{z} = \mathbf{Y}\mathbf{n}_1 + \mathbf{n}_0 \tag{1}$$

where $\mathbf{n}_1 \sim \mathrm{N}_p(\mathbf{0}, \mathbf{R}_1)$ is a p-dimensional normal random vector with zero mean and unknown covariance matrix $\mathbf{R}_1$, and $\mathbf{n}_0 \sim \mathrm{N}_n(\mathbf{0}, \mathbf{R}_0)$ is an n-dimensional normal random vector independent of $\mathbf{n}_1$ with zero mean and known covariance matrix $\mathbf{R}_0$. The normal random vector $\mathbf{z} \sim \mathrm{N}_n(\mathbf{0}, \mathbf{R}_2)$ has zero mean and covariance matrix

$$E[\mathbf{z}\mathbf{z}^{\mathrm{T}}] = \mathbf{R}_2 = \mathbf{Y}\mathbf{R}_1\mathbf{Y}^{\mathrm{T}} + \mathbf{R}_0. \tag{2}$$

Such problems arise, for example, when there is an interference term $\mathbf{Y}\mathbf{n}_1$ with fewer degrees of freedom than the number of available sensors [16], [66]. CRBs for estimation problems in this form are well known [7], [60]. The Fisher information matrix for the unknown parameters $\theta^1, \theta^2, \ldots, \theta^n$ is given by the simple expression $(\mathbf{G})_{ij} = \operatorname{tr}\left(\mathbf{R}_2^{-1}(\partial\mathbf{R}_2/\partial\theta^i)\mathbf{R}_2^{-1}(\partial\mathbf{R}_2/\partial\theta^j)\right)$, and $\mathbf{G}^{-1}$ provides the so-called stochastic CRB. What differs in the estimation problem of (1) from the standard case is an explanation of the derivative terms when the parameters lie on a manifold. The analysis of this problem may be viewed in the context of previous analysis of subspace estimation and superresolution methods [30], [70], [72]–[74], [76], [77]. This paper addresses both the real-valued and (proper [49]) complex-valued cases, which are also referred to as the real symmetric and Hermitian cases, respectively. All real-valued examples may be extended to (proper) in-phase plus quadrature data by replacing transposition with conjugate transposition and using the real representation of the unitary group.

B. Invariance of the Model Estimation Problem

The estimation problem of (1) is invariant to the transformations

$$\mathbf{Y} \mapsto \mathbf{Y}\mathbf{A}^{-1}, \quad \mathbf{R}_1 \mapsto \mathbf{A}\mathbf{R}_1\mathbf{A}^{\mathrm{T}} \tag{3}$$

for any p by p invertible matrix $\mathbf{A}$ in $\boldsymbol{Gl}(p)$, which is the general linear group of real p by p invertible matrices. That is, substituting $\mathbf{Y} := \mathbf{Y}\mathbf{A}^{-1}$ and $\mathbf{n}_1 := \mathbf{A}\mathbf{n}_1$ into (1) leaves the measurement $\mathbf{z}$ unchanged. The only invariant of the transformation $\mathbf{Y} \mapsto \mathbf{Y}\mathbf{A}^{-1}$ is the column span of the matrix $\mathbf{Y}$, and the positive-definite symmetric (Hermitian) structure of covariance matrices is, of course, invariant to the transformation $\mathbf{R}_1 \mapsto \mathbf{A}\mathbf{R}_1\mathbf{A}^{\mathrm{T}}$. Therefore, only the column span of $\mathbf{Y}$ and the covariance matrix of $\mathbf{z}$ may be measured, and we ask how accurately we are able to estimate this subspace, i.e., the column span of $\mathbf{Y}$, in the presence of the unknown covariance matrix $\mathbf{R}_1$.

The parameter space for this estimation problem is the set of all p-dimensional subspaces in $\mathbb{R}^n$, which is known as the Grassmann manifold $G_{n,p}$, and the set of all p by p positive-definite symmetric (Hermitian) matrices P_p, which is the so-called nuisance parameter space. Both $G_{n,p}$ and P_p may be represented by sets of equivalence classes, which are known mathematically as quotient or homogeneous spaces. Although this representation is more abstract, it turns out to be very useful for obtaining closed-form expressions of the necessary geometric objects used in this paper. In fact, both the set of subspaces and the set of covariance matrices are what is known as reductive homogeneous spaces and, therefore, possess natural invariant connections and metrics [10], [12], [25], [32], [38], [50]. Quotient spaces are also the "proper" mathematical description of these manifolds.

A Lie group G is a manifold with differentiable group operations. For $H \subset G$ a (closed) subgroup, the quotient "G/H" denotes the set of equivalence classes $G/H \stackrel{\mathrm{def}}{=} \{\,[g] : g \in G\,\}$, where $[g] \stackrel{\mathrm{def}}{=} gH$ is the equivalence class $gh_1 \equiv gh_2$ for all $h_1, h_2 \in H$. For example, any positive-definitive symmetric matrix $\mathbf{R}$ has the Cholesky decomposition $\mathbf{R} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$, where $\mathbf{A} \in \boldsymbol{Gl}(p)$ (the general linear group) is an invertible matrix with the unique polar decomposition [28] $\mathbf{A} = \mathbf{P}\mathbf{Q}$, where $\mathbf{P} \in P_p$ is a positive-definite symmetric matrix, and

$\mathbf{Q} \in \boldsymbol{O}(p)$ (the orthogonal Lie group) is an orthogonal matrix. Clearly, $\mathbf{R} = \mathbf{P}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{P} = \mathbf{P}^2$ and the orthogonal part $\mathbf{Q}$ of the polar decomposition is arbitrary in the specification of $\mathbf{R}$. Therefore, for any covariance matrix $\mathbf{R} \in \boldsymbol{Gl}(p)$, there is a corresponding equivalence class $[\mathbf{R}^{\frac{1}{2}}] = \mathbf{R}^{\frac{1}{2}} \cdot \boldsymbol{O}(p)$ in the quotient space $\boldsymbol{Gl}(p)/\boldsymbol{O}(p)$, where $\mathbf{R}^{\frac{1}{2}}$ is the unique positive-definitive symmetric square root of $\mathbf{R}$. Thus, this set of covarience matrices P_p may be equated with the quotient space $\boldsymbol{Gl}(p)/\boldsymbol{O}(p)$, allowing application of this space's intrinsic geometric structure to problems involving covariace matrices. In the Hermitian covariance matrix case, the correct identification is $P_p \cong \boldsymbol{Gl}(p, \mathbb{C})/\boldsymbol{U}(p)$, where $\boldsymbol{U}(p)$ is the Lie group of unitary matrices.

Another way of viewing the identification of $P_p \cong \boldsymbol{Gl}(p)/\boldsymbol{O}(p)$ is by the transitive group action [10], [32], [38] seen in (3) of the group $\boldsymbol{Gl}(p)$ acting on P_p via the map $\mathbf{A}: \mathbf{R} \mapsto \mathbf{A}\mathbf{R}\mathbf{A}^{\mathrm{T}}$. This action is almost effective (the matrices $\pm\mathbf{I}$ are the only ones that fix all $\mathbf{R}$, $\pm\mathbf{I}: \mathbf{R} \mapsto = \mathbf{I}\mathbf{R}\mathbf{I}^{\mathrm{T}} = \mathbf{R}$) and has the isotropy (invariance) subgroup of $\boldsymbol{O}(p)$ at $\mathbf{R} = \mathbf{I}$ because $\mathbf{Q}: \mathbf{I} \mapsto \mathbf{Q}\mathbf{I}\mathbf{Q}^{\mathrm{T}} = \mathbf{I}$ for all orthogonal matrices $\mathbf{Q}$. The only part of this group action that matters is the positive-definite symmetric part because $\mathbf{A}: \mathbf{I} \mapsto \mathbf{A}\mathbf{I}\mathbf{A}^{\mathrm{T}} = \mathbf{P}^2 = \mathbf{R}$. Thus, the set of positive-definite symmetric (Hermitian) matrices may be viewed as the equivalence class of invertible matrices multiplied on the right by an arbitrary orthogonal (unitary) matrix.

Although covariance matrices obviously have a unique matrix representation, this is not true of subspaces, because for subspaces, it is only the image (column span) of the matrix that matters. Hence, quotient space methods are essential in the description of problems involving subspaces. Edelman *et al.* [23] provide a convenient computational framework for the Grassmann manifold $G_{n,p}$ involving its quotient space structure. Subspaces are represented by a single (nonunique) n by p matrix with orthonormal columns that itself represents the entire equivalence class of matrices with the same column span. Thus, for the unknown matrix $\mathbf{Y}$, $\mathbf{Y}^{\mathrm{T}}\mathbf{Y} = \mathbf{I}$, $\mathbf{Y}$ may be multiplied on the right by any p by p orthogonal matrix, i.e., $\mathbf{Y}\mathbf{Q}$ for $\mathbf{Q} \in \boldsymbol{O}(p)$, without affecting the results. The Grassmann manifold is represented by the quotient $\boldsymbol{O}(n)/\big(\boldsymbol{O}(n-p) \times \boldsymbol{O}(p)\big)$ because the set of n by p orthonormal matrices $\{\, \mathbf{Y}\mathbf{Q} : \mathbf{Q} \in \boldsymbol{O}(p) \,\}$ is the same equivalence class as the set of n by n orthogonal matrices

$$\left\{ (\mathbf{Y}\,\mathbf{Y}_\perp) \begin{pmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}' \end{pmatrix} : \mathbf{Q} \in \boldsymbol{O}(p), \boldsymbol{Q}' \in \boldsymbol{O}(n-p) \right\}$$

where $\mathbf{Y}_\perp$ is an arbitrary n by $(n-p)$ matrix such that $\mathbf{Y}_\perp^{\mathrm{T}}\mathbf{Y}_\perp = \mathbf{I}$ and $\mathbf{Y}^{\mathrm{T}}\mathbf{Y}_\perp = \mathbf{0}$. Many other signal processing applications also involve the Stiefel manifold $V_{n,p} = \boldsymbol{O}(n)/\boldsymbol{O}(n-p)$ of subspace basis vectors or, equivalently, the set of n by p matrices with orthonormal columns [23]. Another representation of the Grassmann manifold $G_{n,p}$ is the set of equivalence classes $V_{n,p}/\boldsymbol{O}(p)$. Although the approach of this paper immediately applies to the Stiefel manifold, it will not be considered here.

The reductive homogeneous space structure of both P_p and $G_{n,p}$ is exploited extensively in this paper, as are the corresponding invariant affine connections and invariant metrics [viz. (65), (66), (119), and (122)]. Reductive homogeneous spaces and their corresponding natural Riemannian metrics appear frequently in signal processing and other applications [1], [33], [43], [65], e.g., the Stiefel manifold in the singular value and QR-decompositions, but their presence is not widely acknowledged. A homogeneous space G/H is said to be reductive [12], [32], [38] if there exists a decomposition $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$ (direct sum) such that $\mathrm{Ad}_H(\mathfrak{m}) \stackrel{\mathrm{def}}{=} \{\, hAh^{-1} : h \in H, A \in \mathfrak{m} \,\} \subset \mathfrak{m}$, where $\mathfrak{g}$ and $\mathfrak{h}$ are the Lie algebras of G and H, respectively. Given a bilinear form (metric) on $\mathfrak{m}$ (e.g., the trace of $\mathbf{AB}$, $\mathrm{tr}\ \mathbf{AB}$, for symmetric matrices), there corresponds a G-invariant metric and a corresponding G-invariant affine connection on G/H. This is said to be the natural metric on G/H, which essentially corresponds to a restriction of the Killing form on $\mathfrak{g}$ in most applications. For the example of the covariance matrices, $\mathfrak{g} = \mathfrak{gl}(p)$, which is the Lie algebra of p by p matrices [or $\mathfrak{gl}(p, \mathbb{C})$], $\mathfrak{h} = \mathfrak{so}(p)$, which is the sub-Lie algebra of skew-symmetric matrices [or the skew-Hermitian matrices $\mathfrak{u}(n)$], and $\mathfrak{m} = \{\text{symmetric Hermitian matrices}\}$, so that $\mathfrak{gl}(p) = \mathfrak{m} + \mathfrak{so}(p)$ (direct sum). That is, any p by p matrix $\mathbf{A}$ may be expressed as the sum of its symmetric part $(1/2)(\mathbf{A} + \mathbf{A}^{\mathrm{T}})$ and its skew-symmetric part $(1/2)(\mathbf{A} - \mathbf{A}^{\mathrm{T}})$. The symmetric matrices are $\mathrm{Ad}_{\boldsymbol{O}(n)}$-invariant because for any symmetric matrix $\mathbf{S}$ and orthogonal matrix $\mathbf{Q}$, $\mathrm{Ad}_{\mathbf{Q}}(\mathbf{S}) = \mathbf{Q}\mathbf{S}\mathbf{Q}^{-1} = \mathbf{Q}\mathbf{S}\mathbf{Q}^{\mathrm{T}}$, which is also symmetric. Therefore, $P_p \cong \boldsymbol{Gl}(p)/\boldsymbol{O}(p)$ admits a $\boldsymbol{Gl}(p)$-invariant metric and connection corresponding to the bilinear form $\mathrm{tr}\ \mathbf{AB}$ at $\mathbf{R} = \mathbf{I}$, specifically (up to an arbitrary scale factor)

$$g_{\mathbf{R}}(\mathbf{A}, \mathbf{B}) = \mathrm{tr}\ \mathbf{A}\mathbf{R}^{-1}\mathbf{B}\mathbf{R}^{-1} \tag{4}$$

at arbitrary $\mathbf{R} \in P_p$. See Edelman *et al.* [23] for the details of the Grassmann manifold $G_{n,p} \cong \boldsymbol{O}(n)/\big(\boldsymbol{O}(n-p) \times \boldsymbol{O}(p)\big)$ for the subspace example.

Furthermore, the Grassmann manifold $G_{n,p}$ and the positive-definite symmetric matrices P_p with their natural metrics are both Riemannian globally symmetric spaces [10], [25], [31], [38], although, except for closed-form expressions for geodesics, parallel translation, and sectional curvature, the rich geometric properties arising from this symmetric space structure are not used in this paper. As an aside for completeness, $G_{n,p}$ is a compact irreducible symmetric space of type BD I, and $P_p \cong \boldsymbol{Gl}(p)/\boldsymbol{O}(p) \cong \mathbb{R} \times \boldsymbol{Sl}(p)/\boldsymbol{SO}(p)$ is a decomposable symmetric space, where the irreducible noncompact component $\boldsymbol{Sl}(n)/\boldsymbol{SO}(p)$ is of type A I [31, ch. 10, secs. 2.3 and 2.6; ch. 5, prop. 4.2; p. 238]. Elements of P_p decompose naturally into the product of the determinant of the covariance matrix multiplied by a covariance matrix with unity determinant, i.e., $\mathbf{R} = (\det \mathbf{R}) \cdot (\mathbf{R}/\det \mathbf{R})$.

C. Plan of the Paper

The paper's body is organized into three major sections, addressing intrinsic estimation theory, covariance matrix estimation, and subspace and covariance estimation. In Section II, the intrinsic CRB and several of its properties are derived and explained using coordinate-free methods. In Section III, the well-known problem of covariance matrix estimation is analyzed using these intrinsic methods. A closed-form expression for the covariance matrix CRB is derived, and the bias and efficiency of the sample covariance matrix estimator is considered. It is shown that the SCM viewed intrinsically is a biased and inefficient estimator and that the bias term reveals the degraded estimation accuracy that is known to occur at low sample support. Intrinsic CRBs for the subspace plus unknown covariance matrix estimation problem of (1) are computed in closed-form in Section IV and compared with a Monte Carlo simulation. The asymptotic efficiency of the standard subspace estimation approach utilizing the SVD is also analyzed.

II. Intrinsic CRB

An intrinsic version of the CRB is developed in this section. There are abundant treatments of the CRB in its classical form [60], [78], its geometry [61], the case of a constrained parameter space [29], [75], [80], and several generalizations of the CRB to arbitrary Riemannian manifolds [11], [34], [43], [52], [53], [68], [79] and statistical manifolds [5]. It is necessary and desirable for the results in this paper to express the CRB using differential geometric language that is independent of any arbitrary coordinate system and a formula for Fisher information [26], [60], [78] that readily yields CRB results for problems found in signal processing. Other intrinsic derivations of the CRB use different mathematical frameworks (comparison theorems of Riemannian geometry, quantum mechanics) that are not immediately applicable to signal processing and specifically to subspace and covariance estimation theory.

Eight concepts from differential geometry are necessary to define and efficiently compute the CRB: a manifold, its tangent space, the differential of a function, covariant differentiation, Riemannian metrics, geodesic curves, sectional/Riemannian curvature, and the gradient of a function. Working definitions of each of these concepts are provided in footnotes; for complete, technical definitions, see Boothby [10], Helgason [31], Kobayashi and Nomizu [38], or Spivak [69]. In addition, Amari [3]–[6] has significantly extended Rao's [55], [56] original geometric treatment of statistics, and Murray and Rice [48] provide a coordinate-free treatment of these ideas. See Table II in Section IV-F of the paper for a list of differential geometric objects and their more familiar counterparts in Euclidean n-space $\mathbb{R}^n$. Higher order terms containing the manifold's sectional [the higher dimensional generalization of the two-dimensional (2-D) Gaussian curvature] and Riemannian curvature [10], [12], [31], [34], [38], [52], [79], [80] also make their appearance in the intrinsic CRB; however, because the CRB is an asymptotic bound for small errors (high SNR and sample support), these terms are negligible for small errors.

A. Fisher Information Metric

Let $f(\mathbf{z}|\boldsymbol{\theta})$ be the probability density function (pdf) of a vector-valued random variable $\mathbf{z}$ in the sample space $\mathbb{R}^N$, given the parameter $\boldsymbol{\theta}$ in the n-dimensional manifold[1] M (n not necessarily equal to N). The CRB is a consequence of a natural metric structure defined on a statistical model, defined by the parameterized set of probability densities $\{\, f(\mathbf{z}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in M \,\}$. Let

$$\ell(\mathbf{z}|\boldsymbol{\theta}) = \log f(\mathbf{z}|\boldsymbol{\theta}) \tag{5}$$

be the log-likelihood function, and let $\boldsymbol{\Theta}$ and $\boldsymbol{\Omega}$ be two tangent vectors[2] on the parameter space. Define the Fisher information metric[3] (FIM) as

$$g(\boldsymbol{\Theta}, \boldsymbol{\Omega}) = E\left[d\ell(\boldsymbol{\Theta})d\ell(\boldsymbol{\Omega})\right] \tag{6}$$

where $d\ell$ is the differential[4] of the log-likelihood function. We may retrench the dependence on explicit tangent vectors and express the FIM as

$$g = E[d\ell \otimes d\ell] \tag{9}$$

where "$\otimes$" denotes the tensor product. The FIM $\mathbf{G}$ is defined with respect to a particular set of coordinate functions or parameters $(\theta^1, \theta^2, \ldots, \theta^n)$ on M:

$$(\mathbf{G})_{ij} = g\left(\frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \theta^j}\right) = E\left[\frac{\partial \ell}{\partial \theta^i}\frac{\partial \ell}{\partial \theta^j}\right]. \tag{10}$$

Refer to footnote 4 and the "Rosetta Stone" table for an explanation of the notation "$\partial/\partial\theta^i$" for tangent vectors.

Of course, the Fisher information metric is invariant to the change of coordinates $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\theta})$, i.e., $\phi^i = \phi^i(\theta^1, \theta^2, \ldots, \theta^n)$ $(i = 1, \ldots, n)$, because the FIM transforms contravariantly

$$\mathbf{G}_{\boldsymbol{\theta}} = \mathbf{J}_{\boldsymbol{\phi}}^{\mathrm{T}} \mathbf{G}_{\boldsymbol{\phi}} \mathbf{J}_{\boldsymbol{\phi}} \tag{11}$$

where $\mathbf{G}_{\boldsymbol{\theta}}$ and $\mathbf{G}_{\boldsymbol{\phi}}$ are the FIMs with respect to the coordinates specified by their subscripts, and $(\mathbf{J}_{\boldsymbol{\phi}})^i{}_j = \partial\phi^i/\partial\theta^j$ is the Jaco-

[1]A manifold is a space that "looks" like $\mathbb{R}^n$ locally, i.e., M may be parameterized by n coordinate functions $\theta^1, \theta^2, \ldots, \theta^n$ that may be arbitrarily chosen up to obvious consistency conditions. (For example, positions on the globe are mapped by a variety of 2-D coordinate systems: latitude-longitude, Mercator, and so forth.)

[2]The tangent space of a manifold at a point is, nonrigorously, the set of vectors tangent to that point; rigorously, it is necessary to define manifolds and their tangent vectors independently of their embedding in a higher dimension [e.g., as we typically view the 2-sphere embedded in three-dimensional (3-D) space]. The tangent space is a vector space of the same dimension as M. This vector space at $\boldsymbol{\theta} \in M$ is typically denoted by $T_{\boldsymbol{\theta}}M$. The dual space of linear maps from $T_{\boldsymbol{\theta}}M$ to $\mathbb{R}$, called the cotangent space, is denoted $T_{\boldsymbol{\theta}}^*M$. We imagine tangent vectors to be column vectors and cotangent vectors to be row vectors. If we have a mapping $\boldsymbol{\phi} : M \to P$ from one manifold to another, then there exists a natural linear mapping $\boldsymbol{\phi}_* : T_{\boldsymbol{\theta}}M \to T_{\boldsymbol{\phi}}P$ from the tangent space at $\boldsymbol{\theta}$ to the tangent space at $\boldsymbol{\phi}(\boldsymbol{\theta})$. If coordinates are fixed, $\boldsymbol{\phi}_* = \partial\boldsymbol{\phi}/\partial\boldsymbol{\theta}$, i.e., simply the Jacobian transformation. This linear map is called the push-forward: a notion that generalizes Taylor's theorem $\boldsymbol{\phi}(\boldsymbol{\theta} + \delta\boldsymbol{\theta}) = \boldsymbol{\phi}(\boldsymbol{\theta}) + (\partial\boldsymbol{\phi}/\partial\boldsymbol{\theta})\delta\boldsymbol{\theta} +$ h.o.t. $= \boldsymbol{\phi} + \delta\boldsymbol{\phi}$.

[3]A Riemannian metric g is defined by a nondegenerate inner product on the manifold's tangent space, i.e., $g: T_{\boldsymbol{\theta}}M \times T_{\boldsymbol{\theta}}M \to \mathbb{R}$ is a definite quadratic form at each point on the manifold. If $\boldsymbol{\Omega}$ is a tangent vector, then the square length of $\boldsymbol{\Omega}$ is given by $\|\boldsymbol{\Omega}\|^2 = \langle \boldsymbol{\Omega}, \boldsymbol{\Omega}\rangle \overset{\text{def}}{=} g(\boldsymbol{\Omega}, \boldsymbol{\Omega})$. Note that this inner product depends on the location of the tangent vector. For the example of the sphere embedded in Euclidean space, $S^{n-1} = \{\, \boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 1 \,\}$, $g(\boldsymbol{\Omega}, \boldsymbol{\Omega}) = \boldsymbol{\Omega}^{\mathrm{T}}(I - \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{\Omega}$ for tangent vectors $\boldsymbol{\Omega}$ to the sphere at $\boldsymbol{X} \in \mathbb{R}^n$. For covariance matrices, the natural Riemannian metric is provided in (4); for subspaces, the natural metric is given in (118).

[4]The differential of a real-valued function $\ell: M \to \mathbb{R}$ on a manifold, called $d\ell$, simply represents the standard directional derivative. If $c(t)$ $(t \in \mathbb{R})$ is a curve on the manifold, then $\boldsymbol{\Omega} = (d/dt)|_{t=0}c(t)$ is a tangent vector at $c(0)$, and the directional derivative of ℓ in direction $\boldsymbol{\Omega}$ is given by

$$d\ell(\boldsymbol{\Omega}) \overset{\text{def}}{=} \left(\frac{d}{dt}\right)\bigg|_{t=0} \ell\left(c(t)\right). \tag{7}$$

Examining this definition with respect to a particular coordinate system, ℓ may be viewed as a function $\ell(\theta^1, \theta^2, \ldots, \theta^n)$, the curve $c(t)$ is viewed as $(c^1(t), c^2(t), \ldots, c^n(t))^{\mathrm{T}}$, the tangent vector is given by $\boldsymbol{\Omega} = (\Omega^1, \Omega^2, \ldots, \Omega^n)^{\mathrm{T}} = (dc^1/dt, dc^2/dt, \ldots, dc^n/dt)^{\mathrm{T}}$, and the directional derivative is given by the equation

$$d\ell(\boldsymbol{\Omega}) = \sum_k \frac{\partial \ell}{\partial \theta^k}\Omega^k. \tag{8}$$

If we view $\boldsymbol{\Omega}$ as an n by 1 column vector, then $d\ell = (\partial\ell/\partial\theta^1, \partial\ell/\partial\theta^2, \ldots, \partial\ell/\partial\theta^n)$ is a 1 by n row vector. Furthermore, the basis vectors induced by these coordinates for the tangent space $T_{\boldsymbol{\theta}}M$ are called $(\partial/\partial\theta^1), (\partial/\partial\theta^2), \ldots, (\partial/\partial\theta^n)$ because $d\ell(\partial/\partial\theta^k) = \partial\ell/\partial\theta^k$. The dual basis vectors for the cotangent space $T_{\boldsymbol{\theta}}^*M$ are called $d\theta^1, d\theta^2, \ldots, d\theta^n$ because $d\theta^i(\partial/\partial\theta^j) = \delta^i{}_j$ (Kronecker delta). Using the push-forward concept for tangent spaces described in footnote 2, $\ell: M \to \mathbb{R}$, and $\ell_*: T_{\boldsymbol{\theta}}M \to \mathbb{R}$, i.e., $d\ell = \ell_*$, which is consistent with the interpretation of $d\ell$ as a cotangent (row) vector.

bian of the change of variables. Because the Jacobian $\mathbf{J}_{\phi}$ determines how coordinate transformations affect accuracy bounds, it is sometimes called a sensitivity matrix [45], [61].

Although the definition of the FIM employing first derivatives given by (9) is sufficient for computing this metric, in many applications (such as the ones considered in this paper), complicated derivations involving cross terms are oftentimes encountered. (It can be helpful to apply multivariate analysis to the resulting expressions. See, for example, Exercise 2.5 of Fang and Zhang [24]). Derivations of CRBs are typically much simpler if the Fisher information is expressed using the second derivatives[5] of the log-likelihood function. The following theorem provides this convenience for an arbitrary choice of affine connection ∇, which allows a great deal of flexibility (∇ is *not* the gradient operator; see footnotes 5 and 8). The FIM is independent of the arbitrary metric (and its induced Levi–Civita connection) chosen to define the root-mean-square error (RMSE) on the parameter space.

Theorem 1: Let $f(\mathbf{z}|\boldsymbol{\theta})$ be a family of pdfs parameterized by $\boldsymbol{\theta} \in M$, $\ell = \log f$ be the log-likelihood function, and $g = E[d\ell \otimes d\ell]$ be the Fisher information metric. For any affine connection ∇ defined on M

$$g = -E[\nabla^2 \ell]. \tag{16}$$

Proof: The proof is a straightforward application of the lemma: $E[d\ell] = 0$, which follows immediately from differentiating the identity $\int f(\mathbf{z}|\boldsymbol{\theta})\, d\mathbf{z} = 1$ with respect to $\boldsymbol{\theta}$ and observing that $df = f\, d\ell$. Applying this lemma to $\nabla^2\ell$, which is expressed in arbitrary coordinates as in (12), it is seen that $\int (\nabla^2 \ell) f\, d\mathbf{z} = \int (\partial^2 \ell / \partial\theta^i \partial\theta^j) f\, d\mathbf{z} - \sum_k \Gamma_{ij}^k \int (\partial\ell/\partial\theta^k) f\, d\mathbf{z} = -\int (\partial\ell/\partial\theta^i)(\partial\ell/\partial\theta^j) f\, d\mathbf{z}$, the last equality following from integration by parts, as in classical Cramér–Rao analysis. ∎

Theorem 1 allows us to compute the Fisher information on an arbitrary manifold using precisely the same approach as is done in classical Cramér–Rao analysis. To see why this is so, consider the following expression for the Hessian of the log-likelihood function defined on an arbitrary manifold endowed with the affine structure ∇:

$$(\nabla^2 \ell)_{\boldsymbol{\theta}_0}(\boldsymbol{\Omega}, \boldsymbol{\Omega}) = \left.\frac{d^2}{dt^2}\right|_{t=0} \ell\big(\boldsymbol{\theta}(t)\big) \tag{17}$$

where $\boldsymbol{\theta}(t)$ is a geodesic[6] emanating from $\boldsymbol{\theta}_0$ in the direction $\boldsymbol{\Omega}$. Evaluating $\boldsymbol{\theta}(t)$ for small t [12], [69, vol. 2, ch. 4, props. 1 and 6] yields

$$\boldsymbol{\theta}(t) = \boldsymbol{\theta}_0 + t\boldsymbol{\Omega} + O(t^3). \tag{19}$$

This equation should be interpreted to hold for normal coordinates (in the sense of footnote 6) $\boldsymbol{\theta} = (\theta^1, \theta^2, \dots, \theta^n) \in \mathbb{R}^n$ on M. Applying (17) and (19) to compute the Fisher information metric, it is immediately seen that for normal coordinates $\boldsymbol{\theta} = (\theta^1, \theta^2, \dots, \theta^n) \in \mathbb{R}^n$

$$\big(-E[\nabla^2 \ell]\big)_{ij} = -E\frac{\partial^2 \ell}{\partial\theta^i \partial\theta^j} \tag{20}$$

where the second derivatives of the right-hand side of (20) represent the ordinary Hessian matrix of $\ell(\boldsymbol{\theta})$, interpreted to be a scalar function defined on $\mathbb{R}^n$. As with any bilinear form, the

[5] Differentiating real-valued functions defined on manifolds is straightforward because we may subtract the value of the function at one point on the manifold from its value at a different point. Differentiating first derivatives again or tangent vectors, however, is not well defined (without specifying additional information) because a vector tangent at one point is not (necessarily) tangent at another: Subtraction between two different vector spaces is not intrinsically defined. The additional structure required to differentiate vectors is called the affine connection ∇, so-called because it allows one to "connect" different tangent spaces and compare objects defined separately at each point. [This is *not* the gradient operator; see (47) in footnote 8.] The covariant differential of a function ℓ along tangent vector $\boldsymbol{\Omega}$, written $\nabla\ell(\boldsymbol{\Omega})$ or $\nabla_{\boldsymbol{\Omega}}\ell$, is simply $d\ell(\boldsymbol{\Omega})$. In a particular coordinate system, the ith component of $\nabla\ell$ is $(\partial\ell/\partial\theta^i)$. The ijth component of the second covariant differential $\nabla^2\ell$, which is the generalization of the Hessian, is given by

$$(\nabla^2\ell)_{ij} = \frac{\partial^2 \ell}{\partial\theta^i\partial\theta^j} - \sum_k \Gamma_{ij}^k \frac{\partial\ell}{\partial\theta^k} \tag{12}$$

where

$$\Gamma_{ij}^k = \frac{1}{2}\sum_l g^{kl}\left(\frac{\partial g_{il}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^l}\right) \tag{13}$$

are the Christoffel symbols (of the second kind) of the connection defined by $\nabla_{\partial/\partial\theta^i}(\partial/\partial\theta^j) = \sum_k \Gamma_{ij}^k(\partial/\partial\theta^k)$ or defined by the equation for geodesics $\ddot\theta^k + \sum_{i,j}\Gamma_{ij}^k\dot\theta^i\dot\theta^j = 0$ in footnote 6. It is sometimes more convenient to work with Christoffel symbols of the first kind, which are denoted $\Gamma_{ij,k}$, where $\Gamma_{ij}^k = \sum_l g^{kl}\Gamma_{ij,l}$, i.e., the $\Gamma_{ij,l}$ are given by the expression in (13) without the g^{kl} term. For torsion-free connections (the typical case for applications), the Christoffel symbols possess the symmetries $\Gamma_{ij}^k = \Gamma_{ji}^k$ and $\Gamma_{ik,j} + \Gamma_{jk,i} = 0$. It is oftentimes preferable to use matrix notation $\boldsymbol{\Gamma}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\theta}})$ to express the quadratic terms in the geodesic equation [23]. For a sphere embedded in Euclidean space, the Christoffel symbols are given by the coefficients of the quadratic form $\boldsymbol{\Gamma}(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) = \boldsymbol{x}\cdot(\boldsymbol{\Omega}_1^{\mathrm{T}}\boldsymbol{\Omega}_2)$ for tangent vectors $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ to the sphere at $\boldsymbol{X}$ in $\mathbb{R}^n$ [23] [cf. (122)]. The ijth component of the covariant differential of a vector field $\boldsymbol{\Omega}(\boldsymbol{\theta}) \in T_{\boldsymbol{\theta}}M$ is

$$(\nabla\boldsymbol{\Omega})^i{}_j = \frac{\partial\Omega^i}{\partial\theta^j} + \sum_k \Gamma_{jk}^i \Omega^k. \tag{14}$$

Classical index notation is used here and throughout the paper (see, for example, Spivak [69] or McCullagh [44] for a description). A useful way to compute the covariant derivative is $\nabla_{\boldsymbol{\Theta}}\boldsymbol{\Omega} = \dot{\boldsymbol{\Omega}}(0) + \boldsymbol{\Gamma}(\boldsymbol{\Omega}(0), \boldsymbol{\Theta})$ for the vector field $\boldsymbol{\Omega}(t) = \boldsymbol{\Omega}(\exp_{\boldsymbol{\theta}} t\boldsymbol{\Theta})$ (the "$\exp_{\boldsymbol{\theta}}$" notation for geodesics is explained in footnote 6).
In general, the covariant derivative of an arbitrary tensor $\boldsymbol{T}$ on M along the tangent vector $\boldsymbol{\Theta}$ is given by the limit

$$\nabla_{\boldsymbol{\Theta}}\boldsymbol{T} = \lim_{t\to 0}\frac{\tau_t^{-1}\boldsymbol{T}(t) - \boldsymbol{T}(0)}{t} \tag{15}$$

where $\boldsymbol{T}(t)$ denotes the tensor at the point $\exp_{\boldsymbol{\theta}}(t\boldsymbol{\Theta})$, and τ_t denotes parallel translation from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}(t) = \exp_{\boldsymbol{\theta}}(t\boldsymbol{\Theta})$ [10], [31], [38], [69]. Covariant differentiation generalizes the concept of the directional derivative of $\boldsymbol{T}$ along $\boldsymbol{\Theta}$ and encodes all intrinsic notions of curvature on the manifold [69, vol. 2, ch. 4–8]. The parallel translation of a tensor field $\boldsymbol{B}(t) = \tau_t\boldsymbol{B}_0$ satisfies the differential equation $\nabla_{\dot{\boldsymbol{\theta}}}\boldsymbol{B} = 0$.

[6] A geodesic curve $\boldsymbol{\theta}(t)$ on a manifold is any curve that satisfies the differential equation

$$\nabla_{\dot{\boldsymbol{\theta}}}\dot{\boldsymbol{\theta}} = \ddot\theta^k + \sum_{i,j}\Gamma_{ij}^k\dot\theta^i\dot\theta^j = 0; \quad \boldsymbol{\theta}(0) = \boldsymbol{\theta}_0; \quad \dot{\boldsymbol{\theta}}(0) = \boldsymbol{\Omega}. \tag{18}$$

If ∇ is a Riemannian connection, i.e., $\nabla g \equiv 0$ for the arbitrary Riemannian metric g, (18) yields length minimizing curves. The map $\boldsymbol{\Omega} \mapsto \boldsymbol{\theta}(t) = \exp_{\boldsymbol{\theta}_0}(\boldsymbol{\Omega}t)$ is called the exponential map and provides a natural diffeomorphism between the tangent space at $\boldsymbol{\theta}_0$ and a neighborhood of $\boldsymbol{\theta}_0$ on the manifold. Looking ahead, the notation "exp" explains the appearance of matrix exponentials in the equations for geodesics on the space of covariance matrices P_n and the Grassmann manifold $G_{n,p}$ given in (67) and (120). The geodesic $\boldsymbol{\theta}(t)$ is said to emanate from $\boldsymbol{\theta}_0$ in the $\boldsymbol{\Omega}$ direction. By the exponential map, any arbitrary choice of basis of the tangent space at $\boldsymbol{\theta}_0$ yields a set of coordinates on M near $\boldsymbol{\theta}_0$; the coordinates are called *normal* [31], [38], [69]. One important fact that will be used in Sections III and IV to determine covariance and subspace estimation accuracies is that the length of the geodesic curve from $\boldsymbol{\theta}_0$ to $\boldsymbol{\theta}(t)$ is $|t|\cdot\|\boldsymbol{\Omega}\|$. For the sphere embedded in Euclidean space, geodesics are great circles, as can be seen by verifying that the paths $\boldsymbol{x}(t) = \boldsymbol{x}\cos\sigma t + \boldsymbol{\Omega}\sin\sigma t$ [$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Omega} = 0$, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{\Omega} = 1$; cf. (120)] satisfy the differential equation $\ddot{\boldsymbol{x}} + \boldsymbol{\Gamma}(\dot{\boldsymbol{x}}, \dot{\boldsymbol{x}}) = \mathbf{0}$, where the Christoffel symbol $\boldsymbol{\Gamma}$ for the sphere is given in footnote 5.

off-diagonal terms of the FIM may be computed using the standard process of polarization:

$$g(\mathbf{A},\mathbf{B}) = \frac{1}{4}\big(g(\mathbf{A}+\mathbf{B},\mathbf{A}+\mathbf{B}) - g(\mathbf{A}-\mathbf{B},\mathbf{A}-\mathbf{B})\big). \quad (21)$$

Compare this FIM approach to other intrinsic treatments of CRBs that employ the exponential map explicitly [34], [43], [51], [52].

Thus, Fisher information does not depend on the choice of any particular coordinates or any underlying Riemannian metric (or affine connection). We should expect nothing less than this result, as it is true of classical estimation bounds. For example, discounting a scaling factor, measurements in feet versus meters do not affect range estimation bounds.

B. Intrinsic Estimation Theory

Let $f(\mathbf{z}|\boldsymbol{\theta})$ be a parameterized pdf, where $\boldsymbol{\theta}$ takes on values in the manifold M. An estimator $\hat{\boldsymbol{\theta}}(\mathbf{z})$ of $\boldsymbol{\theta}$ is a mapping from the space of random variables to M. Oftentimes, this space of random variables is a product space comprising many snapshots taken from some underlying distribution. Because addition and subtraction are not defined between points on arbitrary manifolds, an affine connection ∇ on M (see footnotes 5 and 6) must be assumed to make sense of the concept of the mean value of $\hat{\boldsymbol{\theta}}$ on M [51]. Associated with this connection is the exponential map $\exp_{\boldsymbol{\theta}}: T_{\boldsymbol{\theta}}M \to M$ from the tangent space at any point $\boldsymbol{\theta}$ to points on M (i.e., geodesics) and its inverse $\exp_{\boldsymbol{\theta}}^{-1}: M \to T_{\boldsymbol{\theta}}M$. Looking ahead, the notation "exp" and "$\exp^{-1}$" is explained by the form of distances and geodesics on the space of covariance matrices P_n and the Grassmann manifold $G_{n,p}$ provided in (65), (67), (69), (119), and (120). If M is connected and complete, the function $\exp_{\boldsymbol{\theta}}$ is onto [31], but its inverse $\exp_{\boldsymbol{\theta}}^{-1}$ may have multiple values (e.g., multiple windings for great-circle geodesics on the sphere). Typically, the tangent vector of shortest length may be chosen. In the case of conjugate points to $\boldsymbol{\theta}$ [12], [31], [38] (where $\exp_{\boldsymbol{\theta}}$ is singular, e.g., antipodes on the sphere), an arbitrary tangent vector may be specified; however, this case is of little practical concern because the set of such points has measure zero, and the CRB itself is an asymptotic bound used for small errors. The manifold P_n has nonpositive curvature [see (28) in footnote 7]; therefore, it has no conjugate points [38, vol. 2, ch. 8, sec. 8], and in fact, $\exp^{-1}$ is uniquely defined by the unique logarithm of positive-definite matrices in this case. For the Grassmann manifold $G_{n,p}$, $\exp^{-1}$ is uniquely defined by the principal value of the arccosine function.

Definition 1: Given an estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta} \in M$, the expectation of the estimator $\hat{\boldsymbol{\theta}}$ with respect to $\boldsymbol{\theta}$, which is denoted $E_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}] \in M$, and the bias vector field $\boldsymbol{b}(\boldsymbol{\theta}) \in T_{\boldsymbol{\theta}}M$ of $\hat{\boldsymbol{\theta}}$ are defined as

$$E_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}] \stackrel{\text{def}}{=} \exp_{\boldsymbol{\theta}} E\Big[\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}\Big] = \exp_{\boldsymbol{\theta}} \int \Big(\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}\Big) f(\mathbf{z}|\boldsymbol{\theta})\, d\mathbf{z} \quad (22)$$

$$\boldsymbol{b}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \exp_{\boldsymbol{\theta}}^{-1} E_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}] = E\Big[\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}\Big]. \quad (23)$$

Fig. 1 illustrates these definitions. Note that unlike the standard expectation operator $E[\cdot]$, $E_{\boldsymbol{\theta}}[\cdot]$ is not (necessarily) linear as it does not operate on linear spaces. These definitions are independent of arbitrary coordinates $(\theta^1, \theta^2, \ldots, \theta^n)$ on M but do depend on the specific choice of affine connection ∇, just as the bias vector in $\mathbb{R}^n$ depends on vector addition. In fact, because for every connection $\bar{\nabla}$ there is a unique torsion-free connection ∇ with the same geodesics [69, vol. 2, ch. 6, cor. 17], the bias vector really depends on the choice of geodesics on M.

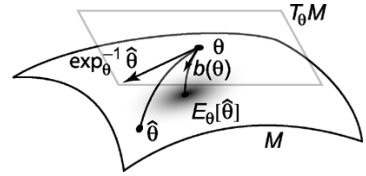

Fig. 1. Intrinsic estimation on a manifold. The estimator $\hat{\boldsymbol{\theta}}(\mathbf{z})$ of the parameter $\boldsymbol{\theta}$ is shown, where $\mathbf{z}$ is taken from the family of pdfs $f(\mathbf{z}|\boldsymbol{\theta})$ whose parameter set is the manifold M. The exponential map "$\exp_{\boldsymbol{\theta}}$" equates points on the manifold with points in the tangent space $T_{\boldsymbol{\theta}}M$ at $\boldsymbol{\theta}$ via geodesics (see footnotes 5 and 6). This structure is necessary to define the expected value of $\hat{\boldsymbol{\theta}}$ because addition and subtraction are not well defined between arbitrary points on a manifold [see (22)]. The bias vector field [see (23)] is defined by $\boldsymbol{b}(\boldsymbol{\theta}) = \exp_{\boldsymbol{\theta}}^{-1} E_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}] = E[\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}]$ and, therefore, depends on the geodesics chosen on M.

Definition 2: The estimator $\hat{\boldsymbol{\theta}}$ is said to be unbiased if $\boldsymbol{b} \equiv 0$ (the zero vector field) so that $E_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$. If $\nabla \boldsymbol{b} \equiv 0$, the bias $\boldsymbol{b}(\boldsymbol{\theta})$ is said to be parallel.

In the ordinary case of Euclidean n-space (i.e., $M = \mathbb{R}^n$), the exponential map and its inverse are simply vector addition, i.e., $\exp_{\boldsymbol{\theta}}\big(\boldsymbol{b}(\boldsymbol{\theta})\big) = \boldsymbol{\theta} + \boldsymbol{b}(\boldsymbol{\theta})$ and $\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$. Parallel bias vectors are constant in Euclidean space because $\nabla \boldsymbol{b} = \partial \boldsymbol{b}/\partial \mathbf{x} \equiv 0$. The proof of the CRB in Euclidean spaces relies on that fact that $(\partial/\partial\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = -\mathbf{I}$. However, for arbitrary Riemannian manifolds, $\nabla \exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}} = -\mathbf{I}$ plus second-order and higher terms involving the manifolds's sectional and Riemannian curvatures.[7] The following lemma quantifies these nonlinear terms, which are negligible for small errors (small distances $d(\hat{\boldsymbol{\theta}}, \exp_{\boldsymbol{\theta}} \boldsymbol{b})$ between $\hat{\boldsymbol{\theta}}$ and $\exp_{\boldsymbol{\theta}} \boldsymbol{b}$) and small biases (small vector norm $\|\boldsymbol{b}\|$), although these higher order terms do appear in the intrinsic CRB.

[7]The sectional curvature is Riemann's higher dimensional generalization of the Gaussian curvature encountered in the study of 2-D surfaces. Given a circle of radius r in a 2-D subspace H of the tangent space $T_{\boldsymbol{\theta}}M$, let $A_0(r) = \pi r^2$ be the area of this circle, and let $A(r)$ be the area of the corresponding "circle" in the manifold mapped using the function $\exp_{\boldsymbol{\theta}}$. Then, the sectional curvature of H is defined [31] to be

$$K(H) = \lim_{r\to 0} 12\,\frac{A_0(r) - A(r)}{r^2 A_0(r)}. \quad (24)$$

We will also write K_M to specify the manifold. For the example of the unit sphere S^{n-1}, $A(r) = 2\pi(1 - \cos r)$, and $K_{S^{n-1}} \equiv 1$, i.e., the unit sphere is a space of constant curvature unity. Equation (24) implies that the area in the manifold is smaller for planes with positive sectional curvature, i.e., geodesics tend to coalesce, and conversely that the area in the manifold is larger for planes with negative sectional curvature, i.e., geodesics tend to diverge from each other. Given tangent vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ and the vector field $\boldsymbol{Z}$, the Riemannian curvature tensor is defined to be

$$\boldsymbol{R}(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z} = \nabla_X \nabla_Y \boldsymbol{Z} - \nabla_Y \nabla_X \boldsymbol{Z} - \nabla_{[X,Y]}\boldsymbol{Z} \quad (25)$$

where $[\boldsymbol{X},\boldsymbol{Y}] = \boldsymbol{XY} - \boldsymbol{YX} \in T_{\boldsymbol{\theta}}M$ is the Lie bracket. The Riemannian curvature $\boldsymbol{R}(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}$ measures the amount by which the vector $\boldsymbol{Z}$ is altered as it is parallel translated around the "parallelogram" in the manifold defined by $\boldsymbol{X}$ and $\boldsymbol{Y}$ [31, ch. 1, ex. E.2]: $\boldsymbol{R}(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z} = \lim_{t\to 0}(\boldsymbol{Z} - \boldsymbol{Z}(t))/t^2$, where $\boldsymbol{Z}(t)$ is the parallel translation of $\boldsymbol{Z}$. Remarkably, sectional curvature and Riemannian curvature are equivalent: The $n(n-1)/2$ sectional curvatures completely determine the Riemannian curvature tensor and vice-versa. The relationship between the two is

$$K(\boldsymbol{X}\wedge\boldsymbol{Y}) = \frac{\langle \boldsymbol{R}(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Y}, \boldsymbol{X}\rangle}{\|\boldsymbol{X}\wedge\boldsymbol{Y}\|^2} \quad (26)$$

where $K(\boldsymbol{X}\wedge\boldsymbol{Y})$ is the sectional curvature of the subspace spanned by $\boldsymbol{X}$ and $\boldsymbol{Y}$, $\|\boldsymbol{X}\wedge\boldsymbol{Y}\|^2=\|\boldsymbol{X}\|^2\|\boldsymbol{Y}\|^2\sin^2\alpha$ is the square area of the parallelogram formed by $\boldsymbol{X}$ and $\boldsymbol{Y}$, and α is the angle between these vectors. Fig. 2 illustrates how curvature causes simple vector addition to fail.

There are relatively simple formulas [12], [13], [31] for the Riemannian curvature tensor in the case of homogeneous spaces encountered in signal processing applications. Specifically, the Lie group $G=\boldsymbol{Gl}(n,\mathbb{C})$ is noncompact, and its Lie algebra $\mathfrak{g}=\mathfrak{gl}(n,\mathbb{C})$ admits a decomposition $\mathfrak{g}=\mathfrak{m}+\mathfrak{h}$ [direct sum, $\mathfrak{m}=$ Hermitian matrices, $\mathfrak{h}=\mathfrak{u}(n)$], such that

$$[\mathfrak{m},\mathfrak{m}]\subset\mathfrak{h},\quad [\mathfrak{m},\mathfrak{h}]\subset\mathfrak{m},\quad [\mathfrak{h},\mathfrak{h}]\subset\mathfrak{h}. \tag{27}$$

The sectional curvature of the symmetric homogeneous space $P_n\cong\boldsymbol{Gl}(n,\mathbb{C})/\boldsymbol{U}(n)$ is given by [13, prop. 3.39], [31, ch. 5, sec. 3]

$$K_{P_n}(\boldsymbol{X}\wedge\boldsymbol{Y})=-\left\|\left[\frac{1}{2}\boldsymbol{X},\frac{1}{2}\boldsymbol{Y}\right]\right\|_{\mathfrak{h}}^2=-\frac{1}{4}\|[\boldsymbol{X},\boldsymbol{Y}]\|^2=\frac{1}{4}\operatorname{tr}([\boldsymbol{X},\boldsymbol{Y}])^2 \tag{28}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are orthonormal Hermitian matrices, and $[\boldsymbol{X},\boldsymbol{Y}]=\boldsymbol{XY}-\boldsymbol{YX}\in\mathfrak{u}(n)$ is the Lie bracket. (The scalings $\frac{1}{2}$ and $\frac{1}{4}$ arise from the fact that the matrix $\mathbf{R}\in P_n$ corresponds to the eqivalence class $\mathbf{R}^{\frac{1}{2}}\cdot H$ in G/H). Note that P_n has nonpositive curvature everywhere. Similarly for the Grassmann manifold, the Lie group $\boldsymbol{O}(n)$ ($\boldsymbol{U}(n)$ in the complex case) is compact, and its Lie algebra admits a comparable decomposition (see Edelman *et al.* [23, sec. 2.3.2] for application-specific details). The sectional curvature of the Grassmann manifold $G_{n,p}\cong\boldsymbol{O}(n)/(\boldsymbol{O}(n-p)\times\boldsymbol{O}(p))$ equals

$$K_{G_{n,p}}(\boldsymbol{X}\wedge\boldsymbol{Y})=\|[\boldsymbol{X},\boldsymbol{Y}]\|_{\mathfrak{h}}^2=-\frac{1}{2}\operatorname{tr}([\boldsymbol{X},\boldsymbol{Y}])^2 \tag{29}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are skew-symmetric matrices of the form $\begin{pmatrix}\mathbf{0} & -\boldsymbol{A}^{\mathrm{T}}\\ \boldsymbol{A} & \mathbf{0}\end{pmatrix}$, and $\boldsymbol{A}$ is an $(n-p)$ by p matrix. [The signs in (28) and (29) are correct because for an arbitrary skew-symmetric matrix $\boldsymbol{\Omega}$, $\|\boldsymbol{\Omega}\|^2=\operatorname{tr}\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{\Omega}=-\operatorname{tr}\boldsymbol{\Omega}^2$.] The sectional curvature of $G_{n,p}$ is non-negative everywhere. For the cases $p=1$ or $p=n-1$, check that (29) is equivalent to the constant curvature $K_{S^{n-1}}\equiv 1$ for the sphere S^{n-1} given above. To bound the terms in the covariance and subspace CRBs corresponding with these curvatures, we note that

$$\max|K_{P_n}|=\frac{1}{4}\quad\text{and}\quad\max K_{G_{n,p}}=1. \tag{30}$$

Lemma 1: Let $f(\mathbf{z}|\boldsymbol{\theta})$ be a family of pdfs parameterized by $\boldsymbol{\theta}\in M$, $\ell=\log f$ be the log-likelihood function, $\bar{g}$ be an arbitrary Riemannian metric (not necessarily the FIM), ∇ be an affine connection on M corresponding to $\bar{g}$, and, for any estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ with bias vector field $\boldsymbol{b}(\boldsymbol{\theta})$, let the matrix $\mathbf{C}$ denote the covariance of $\boldsymbol{X}-\boldsymbol{b}(\boldsymbol{\theta})$, $\boldsymbol{X}=\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}$, all with respect to the arbitrary coordinates $(\theta^1,\theta^2,\ldots,\theta^n)$ near $\boldsymbol{\theta}$ and corresponding tangent vectors $\boldsymbol{E}_i=\partial/\partial\theta^i$, $i=1,\ldots,n$.

1) Then

$$E\left[\left(\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}-\boldsymbol{b}\right)d\ell\right]=\mathbf{I}-\frac{1}{3}\|\boldsymbol{b}\|^2\boldsymbol{K}(\boldsymbol{b})-\frac{1}{3}\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})+\nabla\boldsymbol{b} \tag{31}$$

where the term $-(1/3)\|\boldsymbol{b}\|^2\boldsymbol{K}(\boldsymbol{b})-(1/3)\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ is defined by the expression

$$-E\left[\nabla\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}\right]=\mathbf{I}-\frac{1}{3}\|\boldsymbol{b}\|^2\boldsymbol{K}(\boldsymbol{b})-\frac{1}{3}\boldsymbol{R}_{\mathrm{m}}(\mathbf{C}). \tag{32}$$

2) For sufficiently small bias norm $\|\boldsymbol{b}\|$, the ijth element of $\boldsymbol{K}(\boldsymbol{b})$ is

$$\left(\boldsymbol{K}(\boldsymbol{b})\right)_{ij}=\begin{cases}\sin^2\alpha_i\cdot K(\boldsymbol{b}\wedge\boldsymbol{E}_i)+O\left(\|\boldsymbol{b}\|^3\right), & \text{if } i=j\\ \left[\sin^2\alpha'_{ij}\cdot K(\boldsymbol{b}\wedge(\boldsymbol{E}_i+\boldsymbol{E}_j))-\sin^2\alpha''_{ij}\right.\\ \left.\cdot K(\boldsymbol{b}\wedge(\boldsymbol{E}_i-\boldsymbol{E}_j))\right]+O\left(\|\boldsymbol{b}\|^3\right), & \text{if } i\neq j\end{cases} \tag{33}$$

where $K(H)$ denotes the sectional curvature of the 2-D subspace $H\subset T_{\boldsymbol{\theta}}M$, and α_i, α'_{ij}, and α''_{ij} are the angles between the tangent vector $\boldsymbol{b}$ and $\boldsymbol{E}_i$, $\boldsymbol{E}_i+\boldsymbol{E}_j$, and $\boldsymbol{E}_i-\boldsymbol{E}_j$, respectively.

3) For sufficiently small covariance norm $\lambda_{\max}(\mathbf{C})$, the matrix $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ is given by the quadratic form

$$\langle\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})\boldsymbol{\Omega},\boldsymbol{\Omega}\rangle=E\left[\langle\boldsymbol{R}(\boldsymbol{X}-\boldsymbol{b},\boldsymbol{\Omega})\boldsymbol{\Omega},\boldsymbol{X}-\boldsymbol{b}\rangle\right] \tag{34}$$

where $\boldsymbol{R}(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}$ is the Riemannian curvature tensor. Furthermore, the matrices $\boldsymbol{K}(\boldsymbol{b})$ and $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ are symmetric, and $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ depends linearly on $\mathbf{C}$.

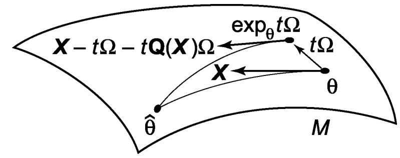

Fig. 2. Failure of vector addition in Riemannian manifolds. Sectional and Riemannian curvature terms (see footnote 7) appearing in the CRB arise from the expression $\nabla\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}$, which contains a quadratic form $\mathbf{Q}(\boldsymbol{X})$ in $\boldsymbol{X}=\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}$. This quadratic form is directly expressible in terms of the Riemannian curvature tensor and the manifold's sectional curvatures [69, vol. 2, ch. 4]. These terms are negligible for the small errors and biases, which are the domain of interest for CRBs, i.e., $d(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})\ll(\max|K_M|)^{-1/2}$, where K_M is the manifold's sectional curvature.

Proof: First take the covariant derivative of the identity $\int(\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}-\boldsymbol{b})f(\mathbf{z}|\boldsymbol{\theta})\,d\mathbf{z}=0$ (the zero vector field). To differentiate the argument inside the integral, recall that for an arbitrary vector field $\boldsymbol{\Omega}$, $\nabla(f\boldsymbol{\Omega})=df\cdot\boldsymbol{\Omega}+f\cdot(\nabla\boldsymbol{\Omega})$. The integral of the first part of this sum is simply the left-hand side of (31); therefore, $E[(\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}-\boldsymbol{b})d\ell]=-E[\nabla(\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}-\boldsymbol{b})]=-E[\nabla\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}]+\nabla\boldsymbol{b}$. This establishes part 1). The remainder of the proof is a straightforward computation using normal coordinates (see footnote 6) and is illustrated in Fig. 2. We will compute $\nabla_{\boldsymbol{\Omega}}\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}$, $\boldsymbol{\Omega}\in T_{\boldsymbol{\theta}}M$, using (15) from footnote 5. Define the tangent vectors $\boldsymbol{X}=\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}\in T_{\boldsymbol{\theta}}M$ and $\boldsymbol{Y}\in T_{\exp_{\boldsymbol{\theta}}t\boldsymbol{\Omega}}M$ by the equations

$$\hat{\boldsymbol{\theta}}=\exp_{\boldsymbol{\theta}}\boldsymbol{X}=\exp_{\exp_{\boldsymbol{\theta}}t\boldsymbol{\Omega}}\boldsymbol{Y}. \tag{35}$$

Assume that the $\{\boldsymbol{E}_i\}$ are an orthonormal basis of $T_{\boldsymbol{\theta}}M$. Expressing all of the quantities in terms of the normal coordinates $\theta^i=\exp_{\boldsymbol{\theta}}\left(\sum_i\theta^i\boldsymbol{E}_i\right)$, the geodesic from $\boldsymbol{\theta}$ to $\hat{\boldsymbol{\theta}}$ is simply the curve $\hat{\boldsymbol{\theta}}=\theta^i(t)=tX^i$, where $\boldsymbol{X}=\sum_i X^i\boldsymbol{E}_i$. The geodesic curve from $\exp_{\boldsymbol{\theta}}t\boldsymbol{\Omega}$ to $\hat{\boldsymbol{\theta}}$ satisfies (18) in footnote 6 subject to the boundary condition specified in (35). Using Riemann's original local analysis of curvature, the metric is expressed as the Taylor series [69, vol. 2, ch. 4]

$$\bar{g}_{ij}=\delta_{ij}+\sum_{kl}a_{ij,kl}\theta^k\theta^l+\cdots \tag{36}$$

where $a_{ij,kl}=(1/2)(\partial^2\bar{g}_{ij}/\partial\theta^k\partial\theta^l)$, which possess the symmetries $a_{ij,kl}=a_{ji,kl}=a_{ij,lk}=a_{kl,ij}$, and $a_{ij,kl}+a_{ik,jl}+a_{il,jk}=0$. Applying (13) in footnote 5 gives the Christoffel symbols (of the first kind) $\Gamma_{ij,k} = -2\sum_l a_{ij,kl}\theta^l$. Solving the geodesic equation for $\boldsymbol{Y}$ to second order in $\boldsymbol{X}$ yields

$$\boldsymbol{Y} = \boldsymbol{X} - t\boldsymbol{\Omega} - t\mathbf{Q}(\boldsymbol{X})\boldsymbol{\Omega} \tag{37}$$

where $\mathbf{Q}(\boldsymbol{X}) = \sum_{kl} a_{ij,kl}X^kX^l$ denotes the quadratic form in (36). The parallel translation of $\boldsymbol{Y}$ along the geodesic $\exp_{\boldsymbol{\theta}} t\boldsymbol{\Omega}$ equals $\boldsymbol{Y}(t) = \boldsymbol{Y} + O(t^2)$. Therefore

$$-\nabla_{\boldsymbol{\Omega}} \exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}} = \boldsymbol{\Omega} + \mathbf{Q}(\boldsymbol{X})\boldsymbol{\Omega} + O\left(\|\boldsymbol{X}\|^3\right) \tag{38}$$

(also see Karcher [37, app. C3]). Riemann's quadratic form $\mathbf{Q}(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{Y}^{\mathrm{T}}\mathbf{Q}(\boldsymbol{X})\boldsymbol{Y} = \sum_{ij,kl} a_{ij,kl}X^kX^lY^iY^j$ simply equals $-(1/3)\|\boldsymbol{X}\wedge\boldsymbol{Y}\|^2K(\boldsymbol{X}\wedge\boldsymbol{Y}) = -(1/3)\langle\boldsymbol{R}(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Y},\boldsymbol{X}\rangle$, where $K(\boldsymbol{X}\wedge\boldsymbol{Y})$ is the sectional curvature of the subspace spanned by $\boldsymbol{X}$, and $\boldsymbol{Y}$, $\|\boldsymbol{X}\wedge\boldsymbol{Y}\|^2 = \|\boldsymbol{X}\|^2\|\boldsymbol{Y}\|^2\sin^2\alpha$ is the square area of the parallelogram formed by $\boldsymbol{X}$ and $\boldsymbol{Y}$, α is the angle between these vectors, and $\boldsymbol{R}(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}$ is the Riemannian curvature tensor (see footnote 7). Taking the inner product of (38) with $\boldsymbol{\Omega}$

$$\begin{aligned}
&\left\langle -\nabla_{\boldsymbol{\Omega}}\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}},\boldsymbol{\Omega}\right\rangle\\
&\quad=\|\boldsymbol{\Omega}\|^2 + \mathbf{Q}(\boldsymbol{X},\boldsymbol{\Omega}) + O\left(\|\boldsymbol{X}\|^3\right)\\
&\quad=\|\boldsymbol{\Omega}\|^2 + \mathbf{Q}\left(\boldsymbol{b}+(\boldsymbol{X}-\boldsymbol{b}),\boldsymbol{\Omega}\right) + O\left(\|\boldsymbol{X}\|^3\right)\\
&\quad=\|\boldsymbol{\Omega}\|^2 + \mathbf{Q}(\boldsymbol{b},\boldsymbol{\Omega}) + \mathbf{Q}(\boldsymbol{X}-\boldsymbol{b},\boldsymbol{\Omega})\\
&\qquad+2\boldsymbol{\Omega}^{\mathrm{T}}\left(\sum_{kl} a_{ij,kl}(\boldsymbol{b})^k(\boldsymbol{X}-\boldsymbol{b})^l\right)\boldsymbol{\Omega} + O\left(\|\boldsymbol{X}\|^3\right)\\
&\quad=\|\boldsymbol{\Omega}\|^2 - \frac{1}{3}\|\boldsymbol{b}\|^2\|\boldsymbol{\Omega}\|^2\sin^2\alpha\cdot K(\boldsymbol{b}\wedge\boldsymbol{\Omega})\\
&\qquad-\frac{1}{3}\langle\boldsymbol{R}(\boldsymbol{X}-\boldsymbol{b},\boldsymbol{\Omega})\boldsymbol{\Omega},\boldsymbol{X}-\boldsymbol{b}\rangle\\
&\qquad+2\boldsymbol{\Omega}^{\mathrm{T}}\left(\sum_{kl} a_{ij,kl}(\boldsymbol{b})^k(\boldsymbol{X}-\boldsymbol{b})^l\right)\boldsymbol{\Omega} + O\left(\|\boldsymbol{X}\|^3\right)
\end{aligned}\tag{39}$$

where α is the angle between $\boldsymbol{\Omega}$ and $\boldsymbol{b}$. Ignoring the higher order terms $O(\|\boldsymbol{X}\|^3)$, the expectation of (39) equals

$$\begin{aligned}
&E\left[\left\langle -\nabla_{\boldsymbol{\Omega}}\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}},\boldsymbol{\Omega}\right\rangle\right]\\
&\quad=\|\boldsymbol{\Omega}\|^2 - \frac{1}{3}\|\boldsymbol{b}\|^2\|\boldsymbol{\Omega}\|^2\sin^2\alpha\cdot K(\boldsymbol{b}\wedge\boldsymbol{\Omega})\\
&\qquad-\frac{1}{3}E\left[\langle\boldsymbol{R}(\boldsymbol{X}-\boldsymbol{b},\boldsymbol{\Omega})\boldsymbol{\Omega},\boldsymbol{X}-\boldsymbol{b}\rangle\right]\\
&\quad=\|\boldsymbol{\Omega}\|^2 - \frac{1}{3}\|\boldsymbol{b}\|^2\|\boldsymbol{\Omega}\|^2\sin^2\alpha\cdot K(\boldsymbol{b}\wedge\boldsymbol{\Omega}) - \frac{1}{3}\boldsymbol{R}_{\mathrm{m}}(\mathbf{C}).
\end{aligned}\tag{40}$$

Applying polarization [(21)] to (40) using the orthonormal basis vectors $\boldsymbol{E}_1,\ldots,\boldsymbol{E}_n$ establishes parts 2) and 3) and shows that $\boldsymbol{K}(\boldsymbol{b})$ and $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ are symmetric. Using normal coordinates, the ijth element of the matrix $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ is seen to be

$$\left(\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})\right)_{ij} = -3\sum_{kl} a_{ij,kl}(\mathbf{C})_{kl} \tag{41}$$

which depends linearly on the covariance matrix $\mathbf{C}$. ∎

As with Gaussian curvature, the units of the sectional and Riemannian curvatures K and $\boldsymbol{R}(\boldsymbol{X},\boldsymbol{Y})$ are the reciprocal of square distance; therefore, the sectional curvature terms in Lemma 1 are negligible for small errors $d(\hat{\boldsymbol{\theta}},\exp_{\boldsymbol{\theta}}\boldsymbol{b})$ and bias norm $\|\boldsymbol{b}\|$ much less than $(\max|K|)^{-1/2}$, i.e., errors and biases less than the reciprocal square root of the maximal sectional curvature. For example, the unit sphere has constant sectional curvature of unity; therefore, the matrices $\boldsymbol{K}(\boldsymbol{b})$ and $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ may be neglected in this case for errors and biases much smaller than 1 rad.

The intrinsic generalization of the Cramér–Rao lower bound is as follows.

Theorem 2 (Cramér–Rao): Let $f(\mathbf{z}|\boldsymbol{\theta})$ be a family of pdfs parameterized by $\boldsymbol{\theta}\in M$, let $\ell = \log f$ be the log-likelihood function, $g = E[d\ell\otimes d\ell]$ be the Fisher information metric, and ∇ be an affine connection on M. 1) For any estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ with bias vector field $\boldsymbol{b}(\boldsymbol{\theta})$, the covariance matrix of $\boldsymbol{X}-\boldsymbol{b}(\boldsymbol{\theta})$, $\boldsymbol{X} = \exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}$ satisfies the matrix inequality

$$\begin{aligned}
\mathbf{C} &+ \frac{1}{3}\left(\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol{b}}^{\mathrm{T}} + \mathbf{M}_{\boldsymbol{b}}\mathbf{G}^{-1}\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})\right)\\
&- \frac{1}{9}\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})\mathbf{G}^{-1}\boldsymbol{R}_{\mathrm{m}}(\mathbf{C}) \geq \mathbf{M}_{\boldsymbol{b}}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol{b}}^{\mathrm{T}}
\end{aligned}\tag{42}$$

where $\mathbf{C} = \mathrm{Cov}(\boldsymbol{X}-\boldsymbol{b}) \stackrel{\mathrm{def}}{=} E[(\boldsymbol{X}-\boldsymbol{b})(\boldsymbol{X}-\boldsymbol{b})^{\mathrm{T}}]$ is the covariance matrix

$$\mathbf{M}_{\boldsymbol{b}} = \mathbf{I} - \frac{1}{3}\|\boldsymbol{b}\|^2\boldsymbol{K}(\boldsymbol{b}) + \nabla\boldsymbol{b} \tag{43}$$

$(\mathbf{G})_{ij} = g(\partial/\partial\theta^i,\partial/\partial\theta^j)$ is the FIM, $\mathbf{I}$ is the identity matrix, $(\nabla\boldsymbol{b})^i{}_j = (\partial\boldsymbol{b}^i/\partial\theta^j) + \sum_k\Gamma^i_{jk}\boldsymbol{b}^k$ is the covariant differential of $\boldsymbol{b}(\boldsymbol{\theta})$, Γ^i_{jk} are the Christoffel symbols, and the matrices $\boldsymbol{K}(\boldsymbol{b})$ and $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ representing sectional and Riemannian curvature terms are defined in Lemma 1, all with respect to the arbitrary coordinates $(\theta^1,\theta^2,\ldots,\theta^n)$ near $\boldsymbol{\theta}$. 2) For $\lambda_{\max}(\mathbf{M}_{\boldsymbol{b}}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol{b}}^{\mathrm{T}})$ sufficiently small relative to $(\max|K_M|)^{-1}$, $\mathbf{C}$ satisfies the matrix inequality

$$\begin{aligned}
\mathbf{C} \geq\; &\mathbf{M}_{\boldsymbol{b}}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol{b}}^{\mathrm{T}}\\
&-\frac{1}{3}\Big(\boldsymbol{R}_{\mathrm{m}}\left(\mathbf{M}_{\boldsymbol{b}}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol{b}}^{\mathrm{T}}\right)\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol{b}}^{\mathrm{T}}\\
&\qquad+\mathbf{M}_{\boldsymbol{b}}\mathbf{G}^{-1}\boldsymbol{R}_{\mathrm{m}}\left(\mathbf{M}_{\boldsymbol{b}}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol{b}}^{\mathrm{T}}\right)^{\mathrm{T}}\Big)\\
&+O\left(\lambda_{\max}\left(\mathbf{M}_{\boldsymbol{b}}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol{b}}^{\mathrm{T}}\right)^3\right).
\end{aligned}\tag{44}$$

3) Neglecting the sectional and Riemannian curvature terms $\boldsymbol{K}(\boldsymbol{b})$ and $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ at small errors and biases, $\mathbf{C}$ satisfies

$$\mathbf{C} \geq (\mathbf{I}+\nabla\boldsymbol{b})\mathbf{G}^{-1}(\mathbf{I}+\nabla\boldsymbol{b})^{\mathrm{T}}. \tag{45}$$

We may substitute "$\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}$" for the expression "$\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}$," interpreting it as the component by component difference for some set of coordinates. In the trivial case $M = \mathbb{R}^n$, the proof of Theorem 2 below is equivalent to asserting that the covariance matrix of the zero-mean random vector

$$\mathbf{v} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} - \boldsymbol{b}(\boldsymbol{\theta}) - \left(\mathbf{I} + \left(\frac{\partial\boldsymbol{b}}{\partial\boldsymbol{\theta}}\right)\right)(\mathrm{grad}_{\boldsymbol{\theta}}\ell) \tag{46}$$

is positive semi-definite, where $\mathrm{grad}_{\boldsymbol\theta}\ell$ is the gradient[8] of ℓ with respect to the FIM $\mathbf{G}$. Readers unfamiliar with technicalities in the proof below are encouraged to prove the theorem in Euclidean space using the fact that $E[\mathbf{v}\mathbf{v}^{\mathrm T}] \geq 0$. As usual, the matrix inequality $\mathbf{A} \geq \mathbf{B}$ is said to hold for positive semi-definite matrices $\mathbf{A}$ and $\mathbf{B}$ if $\mathbf{A}-\mathbf{B} \geq 0$, i.e., $\mathbf{A}-\mathbf{B}$ is positive semi-definite, or $\mathbf{u}^{\mathrm T}(\mathbf{A}-\mathbf{B})\mathbf{u} \geq 0$ for all vectors $\mathbf{u}$.

Proof: The proof is a direct application of Lemma 1 to the computation of the covariance of the random tangent vector

$$\boldsymbol{v} = \exp_{\boldsymbol\theta}^{-1}\hat{\boldsymbol\theta} - \boldsymbol{b}(\boldsymbol\theta) - \left(\mathbf{M}_{\boldsymbol b} - \frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)(\mathrm{grad}_{\boldsymbol\theta}\ell) \quad (51)$$

where $\mathrm{grad}_{\boldsymbol\theta}\ell$ is the gradient of ℓ with respect to the FIM $\mathbf{G}$ (see footnote 8). Denoting $\exp_{\boldsymbol\theta}^{-1}\hat{\boldsymbol\theta}$ by $\boldsymbol{X}$, the covariance of $\boldsymbol{v}$ is given by

$$\begin{aligned}
E[\boldsymbol{v}\boldsymbol{v}^{\mathrm T}] &= E\Bigg[\left(\boldsymbol{X}-\boldsymbol{b}-\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)\mathbf{G}^{-1}d\ell^{\mathrm T}\right)\\
&\qquad\times\left(\boldsymbol{X}-\boldsymbol{b}-\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)\mathbf{G}^{-1}d\ell^{\mathrm T}\right)^{\mathrm T}\Bigg]\\
&= \mathrm{Cov}(\boldsymbol{X}-\boldsymbol{b}) - \left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)\\
&\quad\times\mathbf{G}^{-1}E\left[d\ell^{\mathrm T}(\boldsymbol{X}-\boldsymbol{b})^{\mathrm T}\right]\\
&\quad- E\left[(\boldsymbol{X}-\boldsymbol{b})d\ell\right]\mathbf{G}^{-1}\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)^{\mathrm T}\\
&\quad+\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)\mathbf{G}^{-1}E[d\ell^{\mathrm T}d\ell]\\
&\quad\times\mathbf{G}^{-1}\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)^{\mathrm T}\\
&= \mathbf{C}-\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)\mathbf{G}^{-1}\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)^{\mathrm T}\\
&\quad-\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)\mathbf{G}^{-1}\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)^{\mathrm T}\\
&\quad+\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)\mathbf{G}^{-1}\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)^{\mathrm T}\\
&= \mathbf{C}-\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)\mathbf{G}^{-1}\left(\mathbf{M}_{\boldsymbol b}-\frac{1}{3}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)^{\mathrm T}\\
&= \mathbf{C}-\mathbf{M}_{\boldsymbol b}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol b}^{\mathrm T}\\
&\quad+\frac{1}{3}\left(\boldsymbol{R}_{\mathrm m}(\mathbf{C})\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol b}^{\mathrm T}+\mathbf{M}_{\boldsymbol b}\mathbf{G}^{-1}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\right)\\
&\quad-\frac{1}{9}\boldsymbol{R}_{\mathrm m}(\mathbf{C})\mathbf{G}^{-1}\boldsymbol{R}_{\mathrm m}(\mathbf{C}).
\end{aligned} \quad (52)$$

The mean of $\boldsymbol{v}$ vanishes, $E[\boldsymbol{v}] = \mathbf{o}$, and the covariance of $\boldsymbol{v}$, $E[\boldsymbol{v}\boldsymbol{v}^{\mathrm T}]$ is positive semi-definite, which establishes the first part of the theorem. Expanding $\mathbf{C} = \mathbf{M}_{\boldsymbol b}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol b}^{\mathrm T} + \boldsymbol{\Delta}_1 + \cdots$ into a Taylor series about the zero matrix $\mathbf{0}$ and computing the first-order term $\boldsymbol{\Delta}_1 = -(1/3)\big(\boldsymbol{R}_{\mathrm m}(\mathbf{M}_{\boldsymbol b}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol b}^{\mathrm T})\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol b}^{\mathrm T} + \mathbf{M}_{\boldsymbol b}\mathbf{G}^{-1}\boldsymbol{R}_{\mathrm m}(\mathbf{M}_{\boldsymbol b}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol b}^{\mathrm T})\big)$ establishes the second part. The third part is trivial. ∎

For applications, the intrinsic FIM and CRB are computed as described in Table I. The significance of the sectional and Riemannian curvature terms is an open question that depends upon the specific application; however, as noted earlier, these terms become negligible for small errors and biases. Assuming that the inverse FIM $\mathbf{G}^{-1}$ has units of $(\text{beamwidth})^2/\text{SNR}$ for some beamwidth, as is typical, dimensional analysis of (44) shows that the Riemannian curvature appears in the SNR^{-2} term of the CRB.

Several corollaries follow immediately from Theorem 2. Note that the tensor product $\boldsymbol{v}\otimes\boldsymbol{v}$ of a tangent vector with itself is equivalent to the outer product $\boldsymbol{v}\boldsymbol{v}^{\mathrm T}$ given a particular choice of coordinates [cf. (9) and (10) for cotangent vectors].

Corollary 1: The second-order moment of $\exp_{\boldsymbol\theta}^{-1}\hat{\boldsymbol\theta}$, which is given by

$$E\left[\left(\exp_{\boldsymbol\theta}^{-1}\hat{\boldsymbol\theta}\right)\otimes\left(\exp_{\boldsymbol\theta}^{-1}\hat{\boldsymbol\theta}\right)\right]=\mathrm{Cov}\left(\exp_{\boldsymbol\theta}^{-1}\hat{\boldsymbol\theta}-\boldsymbol{b}(\boldsymbol\theta)\right)+\boldsymbol{b}(\boldsymbol\theta)\otimes\boldsymbol{b}(\boldsymbol\theta) \quad (53)$$

(viewed as a matrix with respect to given coordinates), satisfies the matrix inequality

$$\begin{aligned}
E&\left[\left(\exp_{\boldsymbol\theta}^{-1}\hat{\boldsymbol\theta}\right)\left(\exp_{\boldsymbol\theta}^{-1}\hat{\boldsymbol\theta}\right)^{\mathrm T}\right]\\
&\geq \mathbf{M}_{\boldsymbol b}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol b}^{\mathrm T}+\boldsymbol{b}(\boldsymbol\theta)\boldsymbol{b}(\boldsymbol\theta)^{\mathrm T}\\
&\quad-\frac{1}{3}\Big(\boldsymbol{R}_{\mathrm m}\left(\mathbf{M}_{\boldsymbol b}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol b}^{\mathrm T}\right)\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol b}^{\mathrm T}\\
&\qquad+\mathbf{M}_{\boldsymbol b}\mathbf{G}^{-1}\boldsymbol{R}_{\mathrm m}\left(\mathbf{M}_{\boldsymbol b}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol b}^{\mathrm T}\right)\Big)
\end{aligned} \quad (54)$$

for sufficiently small $\|\boldsymbol{b}\|^2$ and $\lambda_{\max}(\mathbf{M}_{\boldsymbol b}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol b}^{\mathrm T})$ relative to $(\max|K_M|)^{-1}$. Neglecting the sectional and Riemannian curvature terms $\boldsymbol{K}(\boldsymbol{b})$ and $\boldsymbol{R}_{\mathrm m}(\mathbf{C})$ at small errors and biases

$$E\left[\left(\exp_{\boldsymbol\theta}^{-1}\hat{\boldsymbol\theta}\right)\left(\exp_{\boldsymbol\theta}^{-1}\hat{\boldsymbol\theta}\right)^{\mathrm T}\right]\geq(\mathbf{I}+\nabla\boldsymbol{b})\mathbf{G}^{-1}(\mathbf{I}+\nabla\boldsymbol{b})^{\mathrm T}+\boldsymbol{b}(\boldsymbol\theta)\boldsymbol{b}(\boldsymbol\theta)^{\mathrm T}. \quad (55)$$

[8]The gradient of a function ℓ is defined to be the unique tangent vector $\mathrm{grad}\,\ell \in T_{\boldsymbol\theta}M$ such that $g(\mathrm{grad}\,\ell, \boldsymbol\Omega) = d\ell(\boldsymbol\Omega)$ for all tangent vectors $\boldsymbol\Omega$. With respect to particular coordinates

$$\mathrm{grad}\,\ell = \mathbf{G}^{-1}d\ell^{\mathrm T} \quad (47)$$

i.e., the ith element of grad ℓ is $(\mathrm{grad}\,\ell)^i = \sum_j g^{ij}(\partial\ell/\partial\theta^j)$, where g^{ij} is the ijth element of the matrix $\mathbf{G}^{-1}$. Multiplication of any tensor by g^{ij} (summation implied) is called "raising an index," e.g., if $A_j = \partial\ell/\partial\theta^j$ are the coefficients of the differential $d\ell$, then $A^i = \sum_j g^{ij}A_j$ are the coefficients of grad ℓ. The presence of $\mathbf{G}^{-1}$ in the gradient accounts for its appearance in the CRB.
The process of inverting a Riemannian metric is clearly important for computing the CRB. Given the metric $g\colon T_{\boldsymbol\theta}M\times T_{\boldsymbol\theta}M \to \mathbb{R}$, there is a corresponding tensor $g^{-1}\colon T^*_{\boldsymbol\theta}M\times T^*_{\boldsymbol\theta}M \to \mathbb{R}$ naturally defined by the equation

$$g^{-1}(\boldsymbol\omega,\boldsymbol\omega) = g(\boldsymbol\Omega,\boldsymbol\Omega) \quad (48)$$

for all tangent vectors $\boldsymbol\Omega$, where the cotangent vector $\boldsymbol\omega \in T^*_{\boldsymbol\theta}M$ is defined by the equation $g(\boldsymbol\Omega, X) = \boldsymbol\omega(X)$ for all $X \in T_{\boldsymbol\theta}M$ (see footnotes 2 and 4 for the definition of the cotangent space $T^*_{\boldsymbol\theta}M$). The coefficients of the metric g_{ij} and the inverse metric g^{ij} with respect to a specific basis are computed as follows. Given an arbitrary basis $(\partial/\partial\theta^1), (\partial/\partial\theta^2), \ldots, (\partial/\partial\theta^n)$ of the tangent space $T_{\boldsymbol\theta}M$ and the corresponding dual basis $d\theta^1, d\theta^2, \ldots, d\theta^n$ of the cotangent space $T^*_{\boldsymbol\theta}M$ such that $d\theta^i(\partial/\partial\theta^j) = \delta^i{}_j$ (Kronecker delta), we have

$$g_{ij} = g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) \quad (49)$$

$$g^{ij} = g^{-1}(d\theta^i, d\theta^j). \quad (50)$$

Then, $\sum_{k,l} g^{ik}g^{jl}g_{kl} = g^{ij}$ (tautologically, raising both indices of g_{ij}), and the coefficients g^{ij} of the inverse metric express the CRB with respect to this basis.

TABLE I

COMPUTATION OF THE INTRINSIC FIM AND CRB

- Given the log-likelihood function $\ell(\mathbf{z}|\boldsymbol{\theta})$ with $\boldsymbol{\theta} \in M$ and $\boldsymbol{\Omega} \in T_{\boldsymbol{\theta}}M$, compute the Fisher information metric (FIM) using Theorem 1:
$$g_{\rm fim}(\boldsymbol{\Omega},\boldsymbol{\Omega}) = -E\big[(d^2/dt^2)|_{t=0}\ell(\mathbf{z}|\boldsymbol{\theta}+t\boldsymbol{\Omega})\big].$$
This is a quadratic form in $\boldsymbol{\Omega}$.
- Polarize the FIM using the standard formula for quadratic forms [Eq. (20)]:
$$g_{\rm fim}(\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2) = \tfrac{1}{4}\big(g_{\rm fim}(\boldsymbol{\Omega}_1+\boldsymbol{\Omega}_2,\boldsymbol{\Omega}_1+\boldsymbol{\Omega}_2) - g_{\rm fim}(\boldsymbol{\Omega}_1-\boldsymbol{\Omega}_2,\boldsymbol{\Omega}_1-\boldsymbol{\Omega}_2)\big).$$
- Select the desired basis $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, …, $\boldsymbol{\Omega}_n$ for $T_{\boldsymbol{\theta}}M$ and compute the elements g_{ij} of the n-by-n Fisher information matrix $\mathbf{G}$ with respect to this basis as in Eq. (10):
$$g_{ij} = g_{\rm fim}(\boldsymbol{\Omega}_i,\boldsymbol{\Omega}_j).$$
Cramér-Rao bounds derived for this FIM represent accuracy bounds on the coefficients A^i, $i=1,\dots,n$, in the vector decomposition $\boldsymbol{\Omega} = \sum_i A^i\boldsymbol{\Omega}_i$. If $M = \mathbb{R}^n$, this step is equivalent to assigning the matrix elements $g_{ij} = \mathbf{e}_i^{\rm T}\mathbf{G}\mathbf{e}_j$, where $\mathbf{e}_i = (0,\dots,0,1,0,\dots,0)^{\rm T}$ is the i-th standard basis element of $\mathbb{R}^n$.
- To compute the unbiased CRB, invert the FIM using one of the following methods:
  - Invert the matrix $\mathbf{G}$ directly.
  - Compute the (inverse) quadratic form $g_{\rm fim}^{-1}\colon T_{\boldsymbol{\theta}}^*M \times T_{\boldsymbol{\theta}}^*M \to \mathbb{R}$ using Eq. (26) of footnote 7 and the dual basis $\boldsymbol{\Omega}_1^*$, $\boldsymbol{\Omega}_2^*$, …, $\boldsymbol{\Omega}_n^*$ for the cotangent space $T_{\boldsymbol{\theta}}^*M$ such that $\boldsymbol{\Omega}_i^*(\boldsymbol{\Omega}_j) = \delta_{ij}$, then compute $g^{ij} = (\mathbf{G}^{-1})^{ij}$ directly using Eq. (28):
$$g^{ij} = g_{\rm fim}^{-1}(\boldsymbol{\Omega}_i^*,\boldsymbol{\Omega}_j^*).$$

Corollary 2: Assume that $\hat{\boldsymbol{\theta}}$ is an unbiased estimator, i.e., $\boldsymbol{b}(\boldsymbol{\theta}) \equiv 0$ (the zero vector field). Then, the covariance of the estimation error $\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}$ satisfies the inequality

$$\operatorname{Cov}\Big(\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}\Big) \geq \mathbf{G}^{-1} - \frac{1}{3}\left(\boldsymbol{R}_{\rm m}(\mathbf{G}^{-1})\mathbf{G}^{-1} + \mathbf{G}^{-1}\boldsymbol{R}_{\rm m}(\mathbf{G}^{-1})\right). \tag{56}$$

Neglecting the Riemannian curvature terms $\boldsymbol{R}_{\rm m}(\mathbf{C})$ at small errors, the estimation error satisfies

$$\operatorname{Cov}\left(\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}\right) \geq \mathbf{G}^{-1}. \tag{57}$$

Corollary 3 (Efficiency): The estimator $\hat{\boldsymbol{\theta}}$ achieves the Cramér–Rao bound if and only if

$$\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}} - \boldsymbol{b}(\boldsymbol{\theta}) = \left(\mathbf{I} - \frac{1}{3}\|\boldsymbol{b}\|^2\boldsymbol{K}(\boldsymbol{b}) - \frac{1}{3}\boldsymbol{R}_{\rm m}(\mathbf{C}) + \nabla\boldsymbol{b}\right)\operatorname{grad}_{\boldsymbol{\theta}}\ell. \tag{58}$$

If $\hat{\boldsymbol{\theta}}$ is unbiased, then it achieves the CRB if and only if

$$\exp_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}} = \left(\mathbf{I} - \frac{1}{3}\boldsymbol{R}_{\rm m}(\mathbf{C})\right)\operatorname{grad}_{\boldsymbol{\theta}}\ell. \tag{59}$$

Thus the concept of estimator efficiency [26], [60], [78] depends upon the choice of affine connection, or, equivalently, the choice of geodesics on M, as evidenced by the appearance of the ∇ operator and curvature terms in (58).

Corollary 4: The variance of the estimate of the ith coordinate θ^i is given by the ith diagonal element of the matrix $\mathbf{M}_{\boldsymbol{b}}\mathbf{G}^{-1}\mathbf{M}_{\boldsymbol{b}}^{\rm T}$ [plus the first-order term in (44) for larger errors]. If $\hat{\boldsymbol{\theta}}$ is unbiased, then the variance of the estimate of the ith coordinate is given by the ith diagonal element of the inverse of the FIM $\mathbf{G}^{-1}$ (plus the first order term in (44) for larger errors).

Many applications require bounds not on some underlying parameter space but on a mapping $\boldsymbol{\phi}(\boldsymbol{\theta})$ of that space to another manifold. The following theorem provides the generalization of the classical result. The notation $\boldsymbol{\phi}_* = \partial\boldsymbol{\phi}/\partial\boldsymbol{\theta}$ from footnote 2 is used to designate the push-forward of a tangent vector (e.g., the Jacobian matrix of $\boldsymbol{\phi}$ in $\mathbb{R}^n$).

Theorem 3 (Differentiable Mappings): Let f, ℓ, $\boldsymbol{\theta} \in M$, and g be as in Theorem 2, and let $\boldsymbol{\phi}\colon M \to P$ be a differentiable mapping from M to the p-dimensional manifold P. Given arbitrary bases $(\partial/\partial\theta^1), (\partial/\partial\theta^2), \dots, (\partial/\partial\theta^n)$, and $(\partial/\partial\phi^1), (\partial/\partial\phi^2), \dots, (\partial/\partial\phi^p)$ of the tangent spaces $T_{\boldsymbol{\theta}}M$ and $T_{\boldsymbol{\phi}}P$, respectively (or, equivalently, arbitrary coordinates), for any unbiased estimator $\hat{\boldsymbol{\theta}}$ and its mapping $\hat{\boldsymbol{\phi}} = \boldsymbol{\phi}(\hat{\boldsymbol{\theta}})$

$$E\left[\left(\exp_{\boldsymbol{\phi}}^{-1}\hat{\boldsymbol{\phi}}\right)\left(\exp_{\boldsymbol{\phi}}^{-1}\hat{\boldsymbol{\phi}}\right)^{\rm T}\right] \geq \boldsymbol{\phi}_*\mathbf{G}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\phi}_*^{\rm T} \tag{60}$$

$(\mathbf{G}_{\boldsymbol{\theta}})_{ij} = g(\partial/\partial\theta^i, \partial/\partial\theta^j)$ is the FIM, and the push-forward $\boldsymbol{\phi}_* = \partial\boldsymbol{\phi}/\partial\boldsymbol{\theta}$ is the p by n Jacobian matrix with respect to these basis vectors.

Equation (60) in Theorem 3 may be equated with the change-of-variables formula of (11) (after taking a matrix inverse) only when $\boldsymbol{\phi}$ is one-to-one and onto. In general, neither the inverse function $\boldsymbol{\theta}(\boldsymbol{\phi})$ nor its Jacobian $\partial\boldsymbol{\theta}/\partial\boldsymbol{\phi}$ exist.

III. SAMPLE COVARIANCE MATRIX ESTIMATION

In this section, the well-known problem of covariance matrix estimation is considered, utilizing insights from the previous section. The intrinsic geometry of the set of covariance matrices is used to determine the bias and efficiency of the SCM.

Let $\mathbf{Z} = (\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_K)$ be an n by K matrix whose columns are independent and identically distributed (iid) zero-mean complex Gaussian random vectors with covariance matrix $\mathbf{R} \in P_n \cong \boldsymbol{Gl}(n,\mathbb{C})/\boldsymbol{U}(n)$ (also see Diaconis [18, p. 110, ch. 6D and 6E]; the technical details of this identification in the Hermitian case involve the third isomorphism theorem for the normal subgroup of matrices of the form $e^{j\varphi}\mathbf{I}$). The pdf of $\mathbf{Z}$ is $f(\mathbf{Z}|\mathbf{R}) = \left((\pi^n \det\mathbf{R})^{-1}e^{-\operatorname{tr}\hat{\mathbf{R}}\mathbf{R}^{-1}}\right)^K$, where $\hat{\mathbf{R}} = K^{-1}\mathbf{Z}\mathbf{Z}^{\rm H}$ is the SCM. The log-likelihood of this function is (ignoring constants)

$$\ell(\mathbf{Z}|\mathbf{R}) = -K(\operatorname{tr}\,\hat{\mathbf{R}}\mathbf{R}^{-1} + \log\det\mathbf{R}). \tag{61}$$

We wish to compute the CRBs on the covariance matrix $\mathbf{R}$ and examine in which sense the SCM $\hat{\mathbf{R}}$ (the maximum likelihood estimate) achieves these bounds. By Theorem 1, we may extract the second-order terms of $\ell(\mathbf{Z}|\mathbf{R}+\mathbf{D})$ in $\mathbf{D}$ to compute the FIM. These terms follow immediately from the Taylor series

$$\operatorname{tr}(\mathbf{R}+\mathbf{D})^{-1}\mathbf{A} = \operatorname{tr}\,\mathbf{R}^{-1}\mathbf{A} - \operatorname{tr}\mathbf{D}\mathbf{R}^{-1}\mathbf{A}\mathbf{R}^{-1} + \operatorname{tr}\,(\mathbf{D}\mathbf{R}^{-1})^2\mathbf{A}\mathbf{R}^{-1} + \cdots \tag{62}$$

$$\log\det(\mathbf{R}+\mathbf{D}) = \log\det\mathbf{R} + \operatorname{tr}\,\mathbf{D}\mathbf{R}^{-1} - \frac{1}{2}\operatorname{tr}\,(\mathbf{D}\mathbf{R}^{-1})^2 + \cdots \tag{63}$$

where $\mathbf{A}$ and $\mathbf{D}$ are arbitrary Hermitian matrices. It follows that the FIM for $\mathbf{R}$ is given by

$$g_{\text{cov}}(\mathbf{D},\mathbf{D}) = -E[\nabla^2\ell] = K\,\operatorname{tr}\,(\mathbf{D}\mathbf{R}^{-1})^2 \tag{64}$$

that is, the Fisher information metric for Gaussian covariance matrix estimation is simply the natural Riemannian metric on P_n given in (4) (scaled by K, i.e., $g_{\text{cov}} = K\cdot g_{\mathbf{R}}$); this also corresponds to the natural metric on the symmetric cone $\boldsymbol{Gl}(n,\mathbb{C})/\boldsymbol{U}(n)$ [25]. Given the central limit theorem and the invariance of this metric, this result is not too surprising.

A. Natural Covariance Metric CRB

The formula for distances using the natural metric on P_n of (4) is given by the 2-norm of the vector of logarithms of the generalized eigenvalues between two positive-definite matrices, i.e.,

$$d_{\text{cov}}(\mathbf{R}_1,\mathbf{R}_2) = \left(\sum_k (\log\lambda_k)^2\right)^{\frac{1}{2}} \tag{65}$$

where λ_k are the generalized eigenvalues of the pencil $\mathbf{R}_1 - \lambda\mathbf{R}_2$ or, equivalently, $\mathbf{R}_2 - \lambda\mathbf{R}_1$. If multiplied by $10/\log 10$, this distance between two covariance matrices is measured in decibels, i.e., $d_{\text{cov}}(\mathbf{R}_1,\mathbf{R}_2) = \left(\sum_k(10\cdot\log_{10}\lambda_k)^2\right)^{1/2}$ (dB); using Matlab notation, it is expressed as `norm(10*log10(eig(R1,R2)))`. This distance corresponds to the formula for the Fisher information metric for the multivariate normal distribution [57], [64]. The manifold P_n with its natural invariant metric is *not* flat simply because, *inter alia*, it is not a vector space, its Riemannian metric is not constant, its geodesics are not straight lines [(67)], and its Christoffel symbols are nonzero [see (66)].

Geodesics $\mathbf{R}(t) = \exp_{\mathbf{R}}(t\mathbf{D})$ on the covariance matrices P_n corresponding to its natural metric of (4) satisfy the geodesic equation $\ddot{\mathbf{R}} + \boldsymbol{\Gamma}_{\text{cov}}(\dot{\mathbf{R}},\dot{\mathbf{R}}) = \mathbf{0}$, where the Christoffel symbols are given by the quadratic form

$$\boldsymbol{\Gamma}_{\text{cov}}(\mathbf{A},\mathbf{B}) = -\frac{1}{2}(\mathbf{A}\mathbf{R}^{-1}\mathbf{B} + \mathbf{B}\mathbf{R}^{-1}\mathbf{A}) \tag{66}$$

(see footnote 6). A geodesic emanating from $\mathbf{R}$ in the direction $\mathbf{D}\in T_{\mathbf{R}}P_n$ has the form

$$\exp_{\mathbf{R}}(t\mathbf{D}) = \mathbf{R}^{\frac{1}{2}}\exp\left(t\mathbf{R}^{-\frac{1}{2}}\mathbf{D}\mathbf{R}^{-\frac{1}{2}}\right)\mathbf{R}^{\frac{1}{2}} \tag{67}$$

where $\mathbf{R}^{1/2}$ is the unique positive-definite symmetric (Hermitian) square root of $\mathbf{R}$, and "exp" without a subscript denotes the matrix exponential $e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + (t^2/2)\mathbf{A}^2 + \cdots$. This is equivalent to the representation $\mathbf{R}(t) = \mathbf{R}^{1/2}e^{t\mathbf{D}_0}\mathbf{R}^{1/2}$, where $\mathbf{D}_0$ is a tangent vector at $\mathbf{I}$ [$\operatorname{tr}\,\mathbf{D}_0^2 = 1$ for unit vectors] corresponding to $\mathbf{D}$ [$\operatorname{tr}\,(\mathbf{D}\mathbf{R}^{-1})^2 = 1$ for unit vectors] by the coloring transformation

$$\mathbf{D} = \mathbf{R}^{\frac{1}{2}}\mathbf{D}_0\mathbf{R}^{\frac{1}{2}}. \tag{68}$$

The appearance of the matrix exponential in (67), and in (120) of Section IV-C for subspace geodesics, explains the "exp" notation for geodesics described in footnote 6 [see also (69)]. From the expression for geodesics in (67), the inverse exponential map is

$$\exp_{\mathbf{R}}^{-1}\hat{\mathbf{R}} = \mathbf{R}^{\frac{1}{2}}\left(\log\mathbf{R}^{-\frac{1}{2}}\hat{\mathbf{R}}\mathbf{R}^{-\frac{1}{2}}\right)\mathbf{R}^{\frac{1}{2}} \tag{69}$$

(unique matrix square root and logarithm of positive-definite Hermitian matrices). Because Cramér–Rao analysis provides tight estimation bounds at high SNRs, the explicit use of these geodesic formulas is not typically required; however, the metric of (65) corresponding to this formula is useful to measure distances between covariance matrices.

In the simple case of covariance matrices, the FIM and its inverse may be expressed in closed form. To compute CRBs for covariance matrices, the following n^2 orthonormal basis vectors for the tangent space of P_n (Hermitian matrices in this section) at $\mathbf{I}$ are necessary:

$$\mathbf{D}_{ii}^0 = \text{an } n \text{ by } n \text{ symmetric matrix whose } i\text{th diagonal element is unity, zeros elsewhere} \tag{70}$$

$$\mathbf{D}_{ij}^0 = \text{an } n \text{ by } n \text{ symmetric matrix whose } ij\text{th and } ji\text{th elements are both } 2^{-\frac{1}{2}}, \text{ zeros elsewhere } (i<j) \tag{71}$$

$$\mathbf{D}_{ij}^{\text{h}0} = \text{an } n \text{ by } n \text{ Hermitian matrix whose } ij\text{th element is } 2^{-\frac{1}{2}}\sqrt{-1}, \text{ and } ji\text{th element is } -2^{-\frac{1}{2}}\sqrt{-1}, \text{ zeros elsewhere } (i<j). \tag{72}$$

There are n real parameters along the diagonal of $\mathbf{R}$ plus $2\cdot(1/2)n(n-1)$ real parameters in the off-diagonals, for a total of n^2 parameters. For example, the 2 by 2 Hermitian matrix $\begin{pmatrix} a & c+jd \\ c-jd & b\end{pmatrix}$ is decomposed using four orthonormal basis vectors as

$$\begin{pmatrix} a & c+jd \\ c-jd & b\end{pmatrix} = a\begin{pmatrix}1&0\\0&0\end{pmatrix} + b\begin{pmatrix}0&0\\0&1\end{pmatrix} + 2^{\frac{1}{2}}c\begin{pmatrix}0 & 2^{-\frac{1}{2}}\\ 2^{-\frac{1}{2}} & 0\end{pmatrix} + 2^{\frac{1}{2}}d\begin{pmatrix}0 & j2^{-\frac{1}{2}}\\ -j2^{-\frac{1}{2}} & 0\end{pmatrix} \tag{73}$$

and is therefore represented by the real 4-vector $(a, b, 2^{1/2}c, 2^{1/2}d)$. To obtain an orthonormal basis for the tangent space of P_n at $\mathbf{R}$, color the basis vectors by pre- and post-multiplying by $\mathbf{R}^{1/2}$ as in (68), i.e.,

$$\mathbf{D}_{ij}^{\{\text{h}\}} = \mathbf{R}^{\frac{1}{2}}\mathbf{D}_{ij}^{\{\text{h}\}0}\mathbf{R}^{\frac{1}{2}} \tag{74}$$

where the superscript "{h}" denotes a flag for using either the Hermitian basis vectors $\mathbf{D}_{ij}^{\text{h}0}$ or the symmetric basis vectors $\mathbf{D}_{ij}^0$; for notational convenience, $\mathbf{D}_{ii}^{\{\text{h}\}0} \stackrel{\text{def}}{=} \mathbf{D}_{ii}^0$ is implied.

With respect to this basis, (64) yields metric coefficients

$$(\mathbf{G}_{\rm cov})_{\{\rm h\}ij;\{\rm h'\}i'j'} = g_{\rm cov}\left(\mathbf{D}_{ij}^{\{\rm h\}}, \mathbf{D}_{i'j'}^{\{\rm h'\}}\right) = K\delta_{\rm hh'}\delta_{ii'}\delta_{jj'} \tag{75}$$

(Kronecker delta), that is, $\mathbf{G}_{\rm cov} = K\mathbf{I}_{n^2}$.

A closed-form inversion formula for the covariance FIM is easily obtained via (48) in footnote 8. The fact that the Hermitian matrices are self-dual is used, i.e., if $\mathbf{U}$ is a linear mapping from the Hermitian matrices to $\mathbb{R}$, then $\mathbf{U}$ may also be represented as a Hermitian matrix itself via the definition

$$\mathbf{U}(\mathbf{X}) \stackrel{\rm def}{=} \operatorname{tr}\ \mathbf{UX} \tag{76}$$

for all Hermitian matrices $\mathbf{X}$. The inverse of the Fisher information metric (see footnote 8) is defined by

$$g_{\rm cov}^{-1}(\mathbf{U},\mathbf{U}) = g_{\rm cov}(\mathbf{D},\mathbf{D}) \tag{77}$$

for all $\mathbf{D}$, where $\mathbf{D}$ and $\mathbf{U}$ are related by the equation $g_{\rm cov}(\mathbf{D},\mathbf{X}) = \mathbf{U}(\mathbf{X}) = \operatorname{tr}\ \mathbf{UX}$ for all $\mathbf{X}$. Clearly, from (64)

$$\mathbf{U} = K\mathbf{R}^{-1}\mathbf{D}\mathbf{R}^{-1} \tag{78}$$

$$\mathbf{D} = K^{-1}\mathbf{RUR}. \tag{79}$$

Applying (77) and (79) to (64) yields the formula

$$g_{\rm cov}^{-1}(\mathbf{U},\mathbf{U}) = K^{-1}\operatorname{tr}\ \mathbf{RURU}. \tag{80}$$

To compute the coefficients of the inverse FIM with respect to the basis vectors of (74), the dual basis vectors

$$\mathbf{D}_{ij}^{\{\rm h\}*} = \mathbf{R}^{-\frac{1}{2}}\mathbf{D}_{ij}^{\{\rm h\}0}\mathbf{R}^{-\frac{1}{2}} \tag{81}$$

are used; note that $\mathbf{D}_{ij}^{\{\rm h\}*}(\mathbf{D}_{i'j'}^{\{\rm h'\}}) = \operatorname{tr}\ \mathbf{D}_{ij}^{\{\rm h\}*}\mathbf{D}_{i'j'}^{\{\rm h'\}} = \delta_{\rm hh'}\delta_{ii'}\delta_{jj'}$.

We now have sufficient machinery to prove the following theorem.

Theorem 4: The CRB on the natural distance [see (65)] between $\mathbf{R}$ and any unbiased covariance matrix estimator $\hat{\mathbf{R}}$ of $\mathbf{R}$ is

$$\epsilon_{\rm cov} \geq \frac{n}{K^{\frac{1}{2}}} \qquad \text{(Hermitian case)} \tag{82}$$

$$\epsilon_{\rm cov} \geq \frac{(n(n+1))^{\frac{1}{2}}}{K^{\frac{1}{2}}} \qquad \text{(real symmetric case)} \tag{83}$$

where $\epsilon_{\rm cov} = E[d_{\rm cov}^2(\hat{\mathbf{R}},\mathbf{R})]^{1/2}$ is the root mean square error, and the Riemannian curvature term $\boldsymbol{R}_{\rm m}(\mathbf{C})$ has been neglected. To convert this distance to decibels, multiply $\epsilon_{\rm cov}$ by $10/\log 10$.

Proof: The square error of the covariance measured using the natural distance of (65) is

$$\epsilon_{\rm cov}^2 = \sum_i (D^{ii})^2 + 2\sum_{i<j}\left((D^{ij})^2 + \left(D_{\rm h}^{ij}\right)^2\right) \tag{84}$$

where each of the D^{ii} are the coefficients of the basis vectors $\mathbf{D}_{ii}$ in (74), and $2^{1/2}D_{\{\rm h\}}^{ij}$ $(i<j)$ are the coefficients of the orthonormal basis vectors $\mathbf{D}_{ij}^{\{\rm h\}}$ in the vector decomposition

$$\mathbf{D} = \sum_i D^{ii}\mathbf{D}_{ii} + \sum_{i<j} 2^{\frac{1}{2}}\left(D^{ij}\mathbf{D}_{ij} + D_{\rm h}^{ij}\mathbf{D}_{ij}^{\rm h}\right). \tag{85}$$

Equation (84) follows from the consequence of (65) and (67): The natural distance between covariance matrices along geodesics is given by

$$d_{\rm cov}(\mathbf{R}(t),\mathbf{R}) = |t|\left(\operatorname{tr}\ (\mathbf{DR}^{-1})^2\right)^{\frac{1}{2}} \tag{86}$$

which is simply the Frobenius norm of the whitened matrix $t\mathbf{R}^{-1/2}\mathbf{D}\mathbf{R}^{-1/2}$ (see also see footnote 6). Therefore, the CRB of the natural distance is given by

$$\epsilon_{\rm cov}^2 \geq \operatorname{tr}\ \mathbf{G}_{\rm cov}^{-1} \tag{87}$$

where $\mathbf{G}_{\rm cov}$ is computed with respect to the coefficients D^{ii} and $2^{1/2}D_{\{\rm h\}}^{ij}$. Either by (75) $\mathbf{G}_{\rm cov}^{-1} = K^{-1}\mathbf{I}_{n^2}$ or by (80) and (81)

$$\begin{aligned}\left(\mathbf{G}_{\rm cov}^{-1}\right)^{\{\rm h\}ij;\{\rm h\}ij} &= g_{\rm cov}^{-1}\left(\mathbf{D}_{ij}^{\{\rm h\}*}, \mathbf{D}_{ij}^{\{\rm h\}*}\right) \\ &= K^{-1}\operatorname{tr}\left(\mathrm{D}_{ij}^{\{\rm h\}0}\right)^2 = K^{-1}\end{aligned} \tag{88}$$

for $i \leq j$, which establishes the theorem. The real symmetric case follows because $\mathbf{G}_{\rm cov}^{-1} = 2\cdot K^{-1}\mathbf{I}_{(1/2)n(n+1)}$ [n real parameters along the diagonal of $\mathbf{R}$ plus $(1/2)n(n-1)$ real parameters in the off-diagonals; the additional factor of 2 comes from the real Gaussian pdf of (110)]. ∎

B. Flat Covariance Metric CRB

The flat metric on the space covariance matrices expressed using the Frobenius norm is oftentimes encountered:

$$d_{\substack{\rm cov\\ \rm flat}}(\mathbf{R}_1,\mathbf{R}_2) = \left(\operatorname{tr}\ (\mathbf{R}_1-\mathbf{R}_2)^2\right)^{\frac{1}{2}} = \|\mathbf{R}_1-\mathbf{R}_2\|_F. \tag{89}$$

Theorem 5: The CRB on the flat distance [see (89)] between any unbiased covariance matrix estimator $\hat{\mathbf{R}}$ and $\mathbf{R}$ is

$$\epsilon_{\substack{\rm cov\\ \rm flat}} \geq \left(\frac{\sum_i (R^{ii})^2 + 2\sum_{i<j} R^{ii}R^{jj}}{K}\right)^{\frac{1}{2}} \tag{90}$$

(Hermitian case), where $\epsilon_{\substack{\rm cov\\ \rm flat}} = E[d_{\substack{\rm cov\\ \rm flat}}^2(\hat{\mathbf{R}},\mathbf{R})]^{1/2}$ is the root mean square error (flat distance), and R^{ij} denotes the ijth element of $\mathbf{R}$. In the real symmetric case, the scale factor in the numerator of (90) is $2(\sum_{i\leq j}(R^{ij})^2 + \sum_{i<j}R^{ii}R^{jj})$. The flat and natural CRBs of Theorem 4 coincide when $\mathbf{R} = \mathbf{I}$.

Proof: The proof recapitulates that of Theorem 4. For the flat metric of (89), bounds are required on the individual coefficients of the covariance matrix itself, i.e., R^{ii} for the diagonals of $\mathbf{R}$ and $2^{1/2}R_{\{\rm h\}}^{ij}$ for the real and imaginary off-diagonals. Using these parameters

$$\epsilon_{\substack{\rm cov\\ \rm flat}}^2 = \sum_i (R^{ii})^2 + 2\sum_{i<j}\left((R^{ij})^2 + \left(R_{\rm h}^{ij}\right)^2\right) \tag{91}$$

where the covariance $\mathbf{R}$ is decomposed as the linear sum

$$\mathbf{R} = \sum_i R^{ii}\mathbf{D}_{ii}^0 + \sum_{i<j} 2^{\frac{1}{2}}\left(R^{ij}\mathbf{D}_{ij}^0 + R_{\rm h}^{ij}\mathbf{D}_{ij}^{\rm h0}\right). \tag{92}$$

With respect to the orthonormal basis vectors $\mathbf{D}_{ii}^{0}$ and $\mathbf{D}_{ij}^{\{\mathrm{h}\}0}$ of (70)–(72), the coefficients of the inverse FIM $\mathbf{G}_{\substack{\mathrm{cov}\\ \mathrm{flat}}}^{-1}$ are

$$\begin{aligned}\left(\mathbf{G}_{\substack{\mathrm{Cov}\\ \mathrm{flat}}}^{-1}\right)^{\{\mathrm{h}\}ij;\{\mathrm{h}'\}i'j'} &= g_{\mathrm{cov}}^{-1}\left(\mathbf{D}_{ij}^{\{\mathrm{h}\}0}, \mathbf{D}_{i'j'}^{\{\mathrm{h}'\}0}\right)\\ &= K^{-1}\operatorname{tr}\,\mathbf{R}\mathbf{D}_{ij}^{\{\mathrm{h}\}0}\mathbf{R}\mathbf{D}_{i'j'}^{\{\mathrm{h}'\}0}. \end{aligned} \quad (93)$$

The CRB on the flat distance is

$$\epsilon_{\substack{\mathrm{cov}\\ \mathrm{flat}}}^{2} \geq \operatorname{tr}\,\mathbf{G}_{\substack{\mathrm{Cov}\\ \mathrm{flat}}}^{-1}. \quad (94)$$

A straightforward computation shows that

$$\left(\mathbf{G}_{\substack{\mathrm{cov}\\ \mathrm{flat}}}^{-1}\right)^{ii;ii} = (R^{ii})^2 \quad (i=1,\ldots,n) \quad (95)$$

$$\left(\mathbf{G}_{\substack{\mathrm{cov}\\ \mathrm{flat}}}^{-1}\right)^{ij;ij} = R^{ii}R^{jj} + (R^{ij})^2 - \left(R_{\mathrm{h}}^{ij}\right)^2 \quad (i<j) \quad (96)$$

$$\left(\mathbf{G}_{\substack{\mathrm{cov}\\ \mathrm{flat}}}^{-1}\right)^{\mathrm{h},ij;\mathrm{h},ij} = R^{ii}R^{jj} - (R^{ij})^2 + \left(R_{\mathrm{h}}^{ij}\right)^2 \quad (i<j) \quad (97)$$

establishing the theorem upon summing over $i \leq j$. ∎

C. *Efficiency and Bias of the Sample Covariance Matrix*

We now have established the tools with which to examine the intrinsic bias and efficiency of the sample covariance matrix $\hat{\mathbf{R}} = K^{-1}\mathbf{Z}\mathbf{Z}^{\mathrm{H}}$ in the sense of (23) and Corollary 3. Obviously, $E[\hat{\mathbf{R}}] = \mathbf{R}$, but this linear expectation operation means the integral $\int \hat{\mathbf{R}} f(\mathbf{Z}|\mathbf{R})\, d\mathbf{Z}$, which treats the covariance matrices as a convex cone [25] included in the vector space $\mathbb{R}^{n^2}$ ($\mathbb{R}^{(1/2)n(n+1)}$ for the real, symmetric case; a convex cone is a subset of a vector space that is closed under addition and multiplication by positive real numbers). Instead of standard linear expectation valid for vector spaces, the expectation of $\hat{\mathbf{R}}$ is interpreted intrinsically as $E_{\mathbf{R}}[\hat{\mathbf{R}}] = \exp_{\mathbf{R}} \int (\exp_{\mathbf{R}}^{-1}\hat{\mathbf{R}}) f(\mathbf{Z}|\mathbf{R})\, d\mathbf{Z}$ for various choices of geodesics on P_n, as in (22).

First, observe from the first-order terms of (62) and (63) that

$$d\ell(\mathbf{D}) = K(\operatorname{tr}\,\mathbf{D}\mathbf{R}^{-1}\hat{\mathbf{R}}\mathbf{R}^{-1} - \operatorname{tr}\,\mathbf{D}\mathbf{R}^{-1}) \quad (98)$$

establishing that $\hat{\mathbf{R}}$ maximizes the likelihood. Because the SCM $\hat{\mathbf{R}}$ is the maximum likelihood estimate, it is asymptotically efficient [17], [78] and independent of the geodesics chosen for P_n. From the definition of the gradient (see footnote 8), $d\ell(\mathbf{D}) = g_{\mathrm{cov}}(\operatorname{grad}_{\mathbf{R}}\ell, \mathbf{D})$ for all $\mathbf{D} \in T_{\mathbf{R}}P_n$; therefore, with respect to the FIM g_{cov}

$$\operatorname{grad}_{\mathbf{R}}\ell = \hat{\mathbf{R}} - \mathbf{R}. \quad (99)$$

The well-known case of the flat metric/connection on P_n is examined first. Flat distances between $\hat{\mathbf{R}}$ and $\mathbf{R}$ are given by the Frobenius norm $\|\hat{\mathbf{R}} - \mathbf{R}\|_F$ as in (89), and the geodesic between these covariance matrices is $\mathbf{R}(t) = \mathbf{R} + t(\hat{\mathbf{R}} - \mathbf{R})$, $0 \leq t \leq 1$. Obviously for the flat connection, $\exp_{\mathbf{R}}^{-1}\hat{\mathbf{R}} = \hat{\mathbf{R}} - \mathbf{R}$.

Theorem 6: The sample covariance matrix $\hat{\mathbf{R}}$ is an unbiased and efficient estimator of the covariance $\mathbf{R}$ with respect to the flat metric on the space of covariance matrices P_n.

Proof: Trivially, $E_{\mathbf{R}}[\hat{\mathbf{R}}] = \mathbf{R} + E[\hat{\mathbf{R}} - \mathbf{R}] = \mathbf{R}$ for the flat metric; therefore, $\hat{\mathbf{R}}$ is unbiased with respect to this metric. By Corollary 3, an unbiased estimator $\hat{\mathbf{R}}$ is efficient if $\exp_{\mathbf{R}}^{-1}\hat{\mathbf{R}} = \operatorname{grad}_{\mathbf{R}}\ell$, which, by (99), is true for the flat metric. ∎

The flat metric on P_n has several undesirable practical properties. It is well known that the SCM is not a very good estimate of $\mathbf{R}$ for small sample sizes K (which leads to *ad hoc* techniques such as diagonal loading [67] and reference prior methods [81]), but this dependence of estimation quality on sample support is not fully revealed in Theorem 6, which ensures that the SCM always achieves the best accuracy possible. In addition, in many applications treating the space of covariance matrices as a vector space may lead to degraded algorithm performance, especially when a projection onto submanifolds of P_n (structured covariance matrices) is desired [8]. The flat metric has undesirable geometric properties as well. Because the positive-definite Hermitian matrices are convex, every point on straight-line geodesics is also a positive-definite Hermitian matrix. Nevertheless, these paths may not be extended indefinitely to all $t \in \mathbb{R}$. Therefore, the space P_n endowed with the flat metric is not geodesically complete, i.e., it is not a complete metric space. Furthermore, and much more significantly for applications, the flat connection is not invariant to the group action $\mathbf{R} \mapsto \mathbf{A}\mathbf{R}\mathbf{A}^{\mathrm{H}}$ of $\boldsymbol{Gl}(n,\mathbb{C})$ on the positive-definite Hermitian matrices, i.e., $d_{\substack{\mathrm{cov}\\ \mathrm{flat}}}(\mathbf{R}_1, \mathbf{R}_2) \neq d_{\substack{\mathrm{cov}\\ \mathrm{flat}}}(\mathbf{A}\mathbf{R}_1\mathbf{A}^{\mathrm{H}}, \mathbf{A}\mathbf{R}_2\mathbf{A}^{\mathrm{H}})$. Therefore, the CRB depends on the underlying covariance matrix $\mathbf{R}$, as seen in Theorem 5.

In contrast, the natural metric on P_n in (65) has none of these defects and reveals some remarkable properties about the SCM, as well as yielding root mean square errors that are consistent with the dependence on sample size observed in theory and practice (see Fig. 3). The natural distance RMSE varies with the sample size relative to the CRB, unlike the flat distance RMSE, whose corresponding estimator is efficient and, therefore, equals its Cramér–Rao lower bound at all sample sizes. Furthermore, the natural metric is invariant to the group action of $\boldsymbol{Gl}(n,\mathbb{C})$; therefore, it yields bounds that are independent of the underlying covariance matrix $\mathbf{R}$. In addition, P_n endowed with this metric is a geodesically complete space. Because the natural Riemannian metric on P_n [see (65)] has the properties of accounting for the change in estimation quality of the SCM as the sample size varies, being invariant to the group action of $\boldsymbol{Gl}(n,\mathbb{C})$ and therefore independent of $\mathbf{R}$, and yielding a complete metric space for P_n, it is recommended for the analysis of covariance matrix estimation problems.

Theorem 7: The sample covariance matrix estimator $\hat{\mathbf{R}}$ with respect to the natural metric on P_n is biased and not efficient. The bias vector field and expectation of $\hat{\mathbf{R}}$ with respect to $\mathbf{R}$ are

$$\mathbf{B}(\mathbf{R}) = E\left[\exp_{\mathbf{R}}^{-1}\hat{\mathbf{R}}\right] = -\beta(n,K)\mathbf{R} \quad (100)$$

$$E_{\mathbf{R}}[\hat{\mathbf{R}}] = \exp_{\mathbf{R}}\mathbf{B}(\mathbf{R}) = e^{-\beta(n,K)}\mathbf{R} \quad (101)$$

where

$$\begin{aligned}\beta(n,K) = \frac{1}{n}&\left(n\log K + n - \psi(K-n+1) + (K-n+1)\right.\\ &\left.\times\, \psi(K-n+2) + \psi(K+1) - (K+1)\psi(K+2)\right) \end{aligned} \quad (102)$$

(Hermitian case), and $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function. In the real symmetric case, $\beta(n,K) = \log K - \log 2 -$

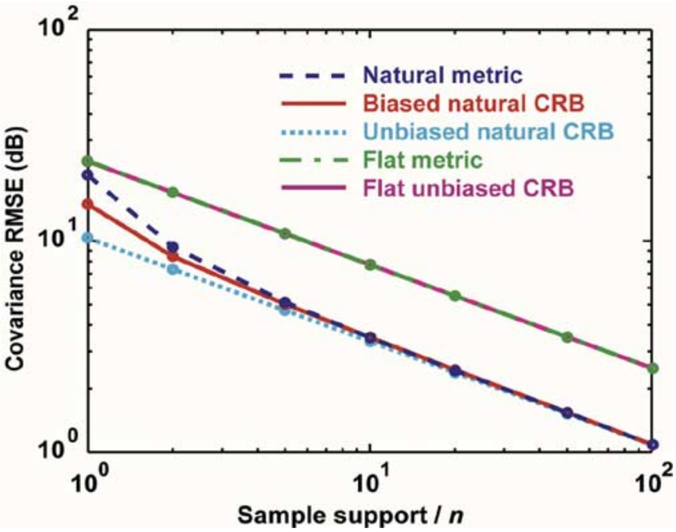

Fig. 3. CRB and RMSE of the distance on P_6 (in decibels, Hermitian case) of the SCM $\hat{\mathbf{R}}$ from $\mathbf{R}$ versus sample support. A Monte Carlo simulation is used with 1000 trials and Wishart distributed SCMs. $\mathbf{R}$ itself is chosen randomly from a Wishart distribution. The unbiased natural CRB [see (82), dotted cyan] is shown below the biased natural CRB [see (105), solid red], which is itself below the unbiased flat CRB [see (90), solid magenta]. The RMSEs of the natural and flat (Frobenius) distances of (65) and (89) are the dashed (blue below green) curves (respectively). Because the SCM w.r.t. the flat metric is an efficient estimator, the RMSE lies on top of its CRB; however, unlike the (inefficient) natural distance, the (efficient) flat distance does not accurately reflect the varying quality of the covariance matrix estimates as the sample size varies, as does the natural distance [Theorem 7 and Corollary 5]. The SCM w.r.t. the natural metric is biased and inefficient and, therefore, does not achieve the biased natural CRB. Nevertheless, the maximum likelihood SCM is always asymptotically efficient. The SCM's Riemannian curvature terms, which become significant for covariance errors on the order of 8.7 dB, have been neglected.

$n^{-1}\sum_{i=1}^{n}\psi\big((1/2)(K-i+1)\big)$. Furthermore, this bias is parallel, and the matrix of sectional curvature terms vanishes, i.e.,

$$\nabla\mathbf{B}\equiv 0 \quad\text{and}\quad \boldsymbol{K}(\mathbf{B})\equiv 0. \tag{103}$$

Proof: From the definition of the bias vector in (23) and the inverse exponential in (69), $\mathbf{B}(\mathbf{R}) = \mathbf{R}^{1/2}E[\log\mathbf{R}^{-1/2}\hat{\mathbf{R}}\mathbf{R}^{-1/2}]\mathbf{R}^{1/2} = \mathbf{R}^{1/2}E[\log\hat{\mathbf{R}}_0]\mathbf{R}^{1/2}$, where $\hat{\mathbf{R}}_0 = \mathbf{R}^{-1/2}\hat{\mathbf{R}}\mathbf{R}^{-1/2}$ is the whitened SCM. Therefore, the bias of the SCM is given by the colored expectation of $\log\hat{\mathbf{R}}_0$. The whitened SCM $\hat{\mathbf{R}}_0$ has the complex Wishart distribution $\mathrm{CW}_K(n, K^{-1}\mathbf{I})$. Using the eigenvalue decomposition of $\hat{\mathbf{R}}_0 = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathrm{H}}$, $E[\log\hat{\mathbf{R}}_0] = E[\mathbf{Q}(\log\boldsymbol{\Lambda})\mathbf{Q}^{\mathrm{H}}] = E[\mathbf{q}\mathbf{q}^{\mathrm{H}}]E[\mathrm{tr}\log\boldsymbol{\Lambda}]$, where $\mathbf{q}$ is an arbitrary column of $\mathbf{Q}$: the last equality following from the independence of eigenvalues and eigenvectors [47]. For the eigenvector part, $E[\mathbf{q}\mathbf{q}^{\mathrm{H}}] = n^{-1}\mathbf{I}$ because $\mathbf{q} = \mathbf{n}/\|\mathbf{n}\|_2$ for a complex normal vector $\mathbf{n}$ (zero mean, unit variance). For the eigenvalue part, $E[\mathrm{tr}\log\boldsymbol{\Lambda}] = E[\log\det\boldsymbol{\Lambda}] = E[\log\det\hat{\mathbf{R}}_0]$, and the distribution of $\det\hat{\mathbf{R}}_0$ is the same as $K^{-n}\prod_{i=1}^{n}\chi^2_{K-i+1}$: the product of independent complex chi-squared random variables [$f_{\chi^2_\nu}(z) = \Gamma(\nu)^{-1}z^{\nu-1}e^{-z}$; Muirhead's Theorem 3.2.15 [47] contains the real case]. Therefore, $E[\log\det\hat{\mathbf{R}}_0] = E[-n\log K + \sum_{i=1}^{n}\log\chi^2_{K-i+1}] = -n\log K + \sum_{i=1}^{n}\psi(K-i+1)$, where $\psi(z)$ is the digamma function. Applying standard identities for $\psi(z)$ evaluated at the integers yields $E[\log\det\hat{\mathbf{R}}_0] = -n\cdot\beta(n,K)$ of (102). Therefore, $E[\log\hat{\mathbf{R}}_0] = -\beta(n,K)\mathbf{I}$, and thus, $\mathbf{B}(\mathbf{R}) = -\beta(n,K)\mathbf{R}$, establishing the first part of the theorem. The proof of the real case is left as an exercise.

To prove that the SCM is not efficient, $\nabla\mathbf{B}$ must be computed. Using the description of the covariant derivative in footnote 5, $(\nabla\mathbf{B})(\mathbf{D}) = \nabla_{\mathbf{D}}\mathbf{B} = \dot{\mathbf{B}}(0) + \boldsymbol{\Gamma}_{\mathrm{cov}}(\mathbf{B}(0),\mathbf{D})$ for any vector field $\mathbf{B}(t) = \mathbf{B}(\exp_{\mathbf{R}} t\mathbf{D})$ along the geodesic $\mathbf{R}(t) = \exp_{\mathbf{R}}(t\mathbf{D})$. For the bias vector field, $\mathbf{B}(t) = -\beta(n,K)\mathbf{R}(t) = -\beta(n,K)\exp_{\mathbf{R}}(t\mathbf{D})$, and $\dot{\mathbf{B}}(0) = -\beta(n,K)\mathbf{D}$. Applying the Christoffel symbols of (66), $\nabla_{\mathbf{D}}\mathbf{B} = \dot{\mathbf{B}}(0) - (1/2)(\mathbf{B}(0)\mathbf{R}^{-1}\mathbf{D} + \mathbf{D}\mathbf{R}^{-1}\mathbf{B}(0)) = -\beta(n,K)\mathbf{D} + (1/2)\beta(n,K)(\mathbf{D}+\mathbf{D}) = 0$. This is true for arbitrary $\mathbf{D}$; therefore, $\nabla\mathbf{B}\equiv 0$, and the bias of the SCM is parallel. This fact can also be shown using the explicit formula for parallel translation on P_n and (15):

$$\tau_t\mathbf{B}_0 = \mathbf{R}^{\frac{1}{2}}e^{\frac{1}{2}t\mathbf{R}^{-\frac{1}{2}}\mathbf{D}\mathbf{R}^{-\frac{1}{2}}}\mathbf{R}^{-\frac{1}{2}}\mathbf{B}_0\mathbf{R}^{-\frac{1}{2}}e^{\frac{1}{2}t\mathbf{R}^{-\frac{1}{2}}\mathbf{D}\mathbf{R}^{-\frac{1}{2}}}\mathbf{R}^{\frac{1}{2}}. \tag{104}$$

The proof that the SCM is not efficient is completed by observing that (58) in Corollary 3 does not hold for $\exp_{\mathbf{R}}^{-1}\hat{\mathbf{R}}$ of (69), $\mathbf{B}(\mathbf{R})$ of (100), and $\mathrm{grad}_{\mathbf{R}}\ell$ of (99). Finally, computing the matrix $\boldsymbol{K}(\mathbf{B})$ defined in (33) at $\mathbf{R}=\mathbf{I}$ gives $\boldsymbol{K}(\mathbf{B}) = \boldsymbol{K}(-\beta(n,K)\mathbf{I}) = \mathbf{0}$, because the formula for the sectional curvature of P_n, $K_{P_n}(\boldsymbol{X}\wedge\boldsymbol{Y}) = -(1/4)\|[\boldsymbol{X},\boldsymbol{Y}]\|^2$ [see (28) of footnote 7] vanishes trivially for $\boldsymbol{X} = -\beta(n,K)\mathbf{I}$. By invariance, $\boldsymbol{K}(\mathbf{B})\equiv\mathbf{0}$ for all $\mathbf{R}$. ∎

It is interesting to note that both $\nabla\mathbf{B}$ and $\boldsymbol{K}(\mathbf{B})$ vanish (conveniently) because the SCMs bias vector is tangent to the 1-D Euclidean part of the symmetric space decomposition [31, ch. 5, sec. 4] $P_n \cong \boldsymbol{Gl}(n)/\boldsymbol{SO}(n) \cong \mathbb{R}\times\boldsymbol{Sl}(n)/\boldsymbol{SO}(n)$, where the $\mathbb{R}$ part represents the determinant of the covariance matrix. That is, the SCMs determinant is biased, but it is otherwise unbiased.

Theorem 7 is of central importance because it quantifies the estimation quality of the sample covariance matrix, especially at low sample support.

Corollary 5: The CRB of the natural distance of the SCM estimate $\hat{\mathbf{R}}$ from $\mathbf{R}$ is

$$\epsilon^2_{\mathrm{cov}} \ge \frac{n^2}{K} + n\beta(n,K)^2 \tag{105}$$

(Hermitian case), where $\epsilon^2_{\mathrm{cov}} = E[d^2_{\mathrm{cov}}(\hat{\mathbf{R}},\mathbf{R})]$ is the mean square error, $\beta(n,K)$ is defined in (102), and the Riemannian curvature term $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ has been neglected.

Proof: The proof is an application of (54) of Corollary 1. Because $\mathbf{B}(\mathbf{R})$ is parallel, the bound on $E[d^2_{\mathrm{cov}}(\hat{\mathbf{R}},\mathbf{R})]$ is given by $\mathrm{tr}\ \mathbf{G}^{-1}_{\mathrm{cov}} + \mathrm{tr}\ \mathbf{B}(\mathbf{R})\otimes\mathbf{B}(\mathbf{R})$, where the tensor product is interpreted to be the outer product over the n^2-dimensional vector space $T_{\mathbf{R}}P_n$ and *not* the Kronecker product of matrices. The first part of this sum is established in (82) of Theorem 4 and the second part by the fact that the trace of the outer product of two vectors equals their inner product: $\mathrm{tr}\ \mathbf{B}(\mathbf{R})\otimes\mathbf{B}(\mathbf{R}) = g_{\mathbf{R}}\big(\mathbf{B}(\mathbf{R}),\mathbf{B}(\mathbf{R})\big) = \mathrm{tr}\ \big(\beta(n,K)^2\mathbf{I}_n\big) = n\beta(n,K)^2$. ∎

An expression for the Riemannian curvature term $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ for the SCM, which becomes significant for errors of order $(10/\log 10)(\max|K_{P_n}|)^{-1/2} = (10/\log 10)2 \approx 8.7$ dB, is possible using the sectional curvature of P_n given in (28) of footnote 7. At $\mathbf{R} = \mathbf{I}$, let $\boldsymbol{X} = \exp_{\mathbf{I}}^{-1} \hat{\mathbf{R}}_0 = \log \hat{\mathbf{R}}_0 \approx \hat{\mathbf{R}}_0 - \mathbf{I}$, and let $\mathbf{D} \in T_{\mathbf{I}} P_n$ be a Hermitian matrix. Then

$$\begin{aligned}\langle \boldsymbol{R}_{\mathrm{m}}(\mathbf{C}) \cdot \mathbf{D}, \mathbf{D} \rangle &= E\left[\langle \boldsymbol{R}(\boldsymbol{X}, \mathbf{D}) \cdot \mathbf{D}, \boldsymbol{X} \rangle\right] \\ &= E\left[\|\boldsymbol{X} \wedge \mathbf{D}\|^2 K_{P_n}(\boldsymbol{X} \wedge \mathbf{D})\right] \\ &= E\left[-\frac{1}{4}\, \|[\boldsymbol{X}, \mathbf{D}]\|^2\right] \\ &= \frac{1}{4} E\left[\operatorname{tr}\left([\hat{\mathbf{R}}_0, \mathbf{D}]\right)^2\right] \end{aligned} \tag{106}$$

where $[\boldsymbol{X}, \boldsymbol{Y}] = \boldsymbol{XY} - \boldsymbol{YX}$ is the Lie bracket. It can be shown that $\langle \mathbf{R}(\mathbf{C}) \cdot \mathbf{D}_{ij}^{\{h\}0}, \mathbf{D}_{ij}^{\{h\}0} \rangle = -(1/2)(n-1)/K$ if $i = j$ and $-(1/2)n/K$ if $i \neq j$.

A connection can also be made with asymptotic results for the eigenvalue spectrum of the SCM. The largest and smallest eigenvalues $\hat{\lambda}_{\max}$ and $\hat{\lambda}_{\min}$ of the whitened SCM $\hat{\mathbf{R}}_0$ are approximated by

$$\hat{\lambda}_{\min,\max} \approx \left(1 \pm \sqrt{\frac{n}{K}}\right)^2 \tag{107}$$

(n large, $K \geq n$) [36], [62]. For $K \approx n$, the extreme spectral values of $\hat{\mathbf{R}}_0$ differ considerably from unity, even though $E[\hat{\mathbf{R}}_0] = \mathbf{I}$ for all K, and this estimator is efficient with respect to the flat connection. Indeed, by the deformed quarter-circle law [36], [42], [63] for the asymptotic distribution of the spectrum of $\hat{\mathbf{R}}_0$ (large n), the mean square of the natural distance of $\hat{\mathbf{R}}_0$ from $\mathbf{I}$ is

$$\begin{aligned} E\left[n^{-1} d_{\mathrm{cov}}^2(\hat{\mathbf{R}}_0, \mathbf{I})\right] &\sim \frac{1}{2\pi y} \int_{(1-\sqrt{y})^2}^{(1+\sqrt{y})^2} (\log \lambda)^2 \\ &\quad \times \frac{\left((1+\sqrt{y})^2 - \lambda\right)^{\frac{1}{2}} \left(\lambda - (1-\sqrt{y})^2\right)^{\frac{1}{2}}}{\lambda}\, d\lambda \\ &= y + \frac{5}{6} y^2 + O(y^3) \end{aligned} \tag{108}$$

($n \to \infty$), where $y = n/K \leq 1$. For large n and large sample support (i.e., small y), the SCMs (biased) CRB in (105) has the asymptotic expansion

$$n^{-1} \epsilon_{\mathrm{cov}}^2 \sim y + \frac{1}{4} y^2 + O(y^3) \quad (n \to \infty) \tag{109}$$

whose linear term in y coincides exactly with that of (108) because the SCM is asymptotically efficient. The SCM is not efficient for $y > 0$ (finite sample support), and the quadratic terms of (108) and (109) differ, with the CRB's quadratic term $(1/4)y^2$ being strictly less than the mean square error's quadratic term $(5/6)y^2$; adding the SCM's Riemannian curvature terms from (108) adds a term of $(1/3)(n^2 - 1)/K^2 \sim (1/3)y^2$, resulting in $(7/12)y^2 < (5/6)y^2$. We note in passing the similarity between the first-order term in $y = n/K$ of this CRB and the well-known Reed–Mallett–Brennan estimation loss for adaptive detectors [58], [67].

IV. Subspace and Covariance Estimation Accuracy

We now return to the analysis of estimating an unknown subspace with an underlying unknown covariance matrix given in (1). The parameter space is the product manifold $M = P_p \times G_{n,p}$, with the covariance matrix being an unknown nuisance parameter. The dimension of this product manifold is the sum of the dimensions of P_p and $G_{n,p}$, which equals $(1/2)p(p+1) + p(n-p)$ in the real case and $p^2 + 2p(n-p)$ in the proper complex case. We will focus on the real case in this section; the proper complex case is a trivial extension.

A. Subspace and Covariance Probability Model

Let $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_K)$ be a real n by K matrix whose columns are iid random vectors defined by (1). The joint pdf of this data matrix is

$$f(\mathbf{Z}|\mathbf{R}_1, \mathbf{Y}) = \left(\frac{e^{-\frac{1}{2} \operatorname{tr} \mathbf{R}_2^{-1} \hat{\mathbf{R}}}}{\left((2\pi)^n \det \mathbf{R}_2\right)^{\frac{1}{2}}}\right)^K \tag{110}$$

$$\mathbf{R}_2 = \mathbf{Y} \mathbf{R}_1 \mathbf{Y}^{\mathrm{T}} + \mathbf{R}_0 \tag{111}$$

where $\hat{\mathbf{R}} = K^{-1} \mathbf{Z}\mathbf{Z}^{\mathrm{T}}$ is the sample covariance matrix, and the n by p and p by p matrices $\mathbf{Y}$ and $\mathbf{R}_1$ represent the unknown subspace and covariance, respectively. The log-likelihood of this function is (ignoring constants)

$$\ell(\mathbf{Z}|\mathbf{R}_1, \mathbf{Y}) = -\frac{1}{2} K \left(\operatorname{tr} \mathbf{R}_2^{-1} \hat{\mathbf{R}} + \log \det \mathbf{R}_2\right). \tag{112}$$

The maximum likelihood estimate of the subspace is simply the p-dimensional principal invariant subspace of the whitened SCM $\hat{\mathbf{R}}_0 = \mathbf{R}_0^{-1/2} \hat{\mathbf{R}} \mathbf{R}_0^{-1/2}$. Indeed, a straightforward computation involving the first-order terms of (62) and (63) establishes that solving the equation $\partial \ell / \partial \mathbf{Y} = \mathbf{0}$ results in the invariant subspace equation [23]

$$\hat{\mathbf{R}}_0 \tilde{\mathbf{Y}} - \tilde{\mathbf{Y}} (\tilde{\mathbf{Y}}^{\mathrm{T}} \hat{\mathbf{R}}_0 \tilde{\mathbf{Y}}) = \mathbf{0}$$

for $\tilde{\mathbf{Y}}$, such that $\tilde{\mathbf{Y}}^{\mathrm{T}} \tilde{\mathbf{Y}} = \mathbf{I}$. Choosing $\tilde{\mathbf{Y}}$ to be the invariant subspace corresponding to the largest eigenvalues of $\hat{\mathbf{R}}_0$ maximizes the log-likelihood.

B. Natural Geometry of Subspaces

Closed-form expressions for CRBs on the parameters $\mathbf{R}_1$ and $\mathbf{Y}$ are obtained if points on the Grassmann manifold are represented by n by p matrices $\mathbf{Y}$ such that

$$\mathbf{Y}^{\mathrm{T}} \mathbf{R}_0^{-1} \mathbf{Y} = \mathbf{I} \tag{113}$$

and post-multiplication of any such matrix by an orthogonal p by p matrix represents the same point in the equivalence class. This constraint is a colored version of the convenient representation of points on $G_{n,p}$ by matrices with orthonormal columns [1], [23] and simply amounts to the whitening transformation

$$\tilde{\mathbf{Y}} = \mathbf{R}_0^{-\frac{1}{2}} \mathbf{Y} \tag{114}$$

where $\tilde{\mathbf{Y}}^{\mathrm{T}} \tilde{\mathbf{Y}} = \mathbf{I}$. Given such a matrix $\tilde{\mathbf{Y}}$ with orthonormal columns or, equivalently, $\mathbf{Y} = \mathbf{R}_0^{1/2} \tilde{\mathbf{Y}}$ satisfying (113), tangent vectors to the Grassmann manifold are represented by n by p matrices $\tilde{\mathbf{\Delta}} \in T_{\tilde{\mathbf{Y}}}G_{n,p}$ or $\mathbf{\Delta} = \mathbf{R}_0^{1/2}\tilde{\mathbf{\Delta}}$ (colored case) of the form

$$\tilde{\mathbf{\Delta}} = \tilde{\mathbf{Y}}_\perp \mathbf{B} \qquad \text{(whitened case)} \tag{115}$$

$$\mathbf{\Delta} = \mathbf{Y}_\perp \mathbf{B} = \mathbf{R}_0^{\frac{1}{2}}\tilde{\mathbf{Y}}_\perp \mathbf{B} \qquad \text{(colored case)} \tag{116}$$

such that $\tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{\Delta}} = \mathbf{0}$ and $\mathbf{Y}^{\mathrm{T}}\mathbf{R}_0^{-1}\mathbf{\Delta} = \mathbf{0}$ or, equivalently, that $\tilde{\mathbf{Y}}_\perp$ is an arbitrary n by $(n-p)$ matrix with orthonormal columns such that $\tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{Y}}_\perp = \mathbf{0}$ and $\tilde{\mathbf{Y}}_\perp^{\mathrm{T}}\tilde{\mathbf{Y}}_\perp = \mathbf{I}$, $\mathbf{Y}_\perp = \mathbf{R}_0^{1/2}\tilde{\mathbf{Y}}_\perp$, and $\mathbf{B}$ is an arbitrary $(n-p)$ by p matrix [23]. Cotangent vectors in $T^*_{\tilde{\mathbf{Y}}}G_{n,p}$ (whitened case) are also represented by n by p matrices $\tilde{\mathbf{\Upsilon}} = \tilde{\mathbf{Y}}_\perp \mathbf{B}$ with the identification

$$\tilde{\mathbf{\Upsilon}}(\tilde{\mathbf{\Delta}}) \stackrel{\text{def}}{=} \operatorname{tr}\, \tilde{\mathbf{\Upsilon}}^{\mathrm{T}}\tilde{\mathbf{\Delta}} \tag{117}$$

for all $\tilde{\mathbf{\Delta}}$ such that $\tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{\Delta}} = \mathbf{0}$. In the colored case, $\mathbf{\Upsilon}(\mathbf{\Delta}) = \tilde{\mathbf{\Upsilon}}(\tilde{\mathbf{\Delta}}) = \operatorname{tr}\, \mathbf{\Upsilon}^{\mathrm{T}}\mathbf{\Delta} = \operatorname{tr}\, \tilde{\mathbf{\Upsilon}}^{\mathrm{T}}\tilde{\mathbf{\Delta}}$; therefore, $\mathbf{\Upsilon} = \mathbf{R}_0^{-1/2}\tilde{\mathbf{\Upsilon}} = \mathbf{R}_0^{-1/2}\tilde{\mathbf{Y}}_\perp\mathbf{B}$, and $\mathbf{Y}^{\mathrm{T}}\mathbf{\Upsilon} = \mathbf{0}$. As usual, dual vectors are whitened contravariantly.

The CRB for the unknown subspace will be computed using the natural metric on $G_{n,p}$ given by [23]

$$g_{\mathrm{ss}}(\tilde{\mathbf{\Delta}}, \tilde{\mathbf{\Delta}}) = \operatorname{tr}\, \tilde{\mathbf{\Delta}}^{\mathrm{T}}(\mathbf{I} - \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathrm{T}})\tilde{\mathbf{\Delta}} = \operatorname{tr}\, \mathbf{\Delta}^{\mathrm{T}}\mathbf{R}_0^{-1}\mathbf{\Delta}. \tag{118}$$

Distances corresponding to this metric are give by the 2-norm of the vector of principal angles between two subspaces, i.e.,

$$d_{\mathrm{ss}}(\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2) = \left(\sum_k \theta_k^2\right)^{\frac{1}{2}} \tag{119}$$

$0 \le \theta_k \le (\pi/2)$, where $\cos\theta_k$ are the singular values of the matrix $\tilde{\mathbf{Y}}_1^{\mathrm{T}}\tilde{\mathbf{Y}}_2$. This distance between two subspaces is measured in radians; in Matlab notation, it is expressed as `norm(acos(svd(orth(Y1)'*orth(Y2))))`. There are several other subspace distances that are defined by embeddings of the Grassmann manifold in higher dimensional manifolds [23, p. 337]. Both the arccosine implicit in (119) and the logarithm for the natural metric on P_p in (65) correspond to the inverse of the exponential map "$\exp^{-1}$" discussed in Section II-B.

Applying (115) to (118) shows that the natural distance between two nearby subspaces is given by the Frobenius norm of the matrix $\mathbf{B}$ ($\|\mathbf{B}\|_F^2 = \operatorname{tr}\, \mathbf{B}^{\mathrm{T}}\mathbf{B}$). Therefore, the CRB of the natural subspace distance is given by the FIM with respect to the elements of $\mathbf{B}$. This fact is made rigorous by the following observations. Geodesics on the Grassmann manifold $G_{n,p}$ corresponding to its natural metric are given by the formula [23]

$$\tilde{\mathbf{Y}}(t) = \tilde{\mathbf{Y}}\mathbf{V}\cos(\mathbf{\Sigma}t)\mathbf{V}^{\mathrm{T}} + \mathbf{U}\sin(\mathbf{\Sigma}t)\mathbf{V}^{\mathrm{T}} \tag{120}$$

where $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}} := \tilde{\mathbf{\Delta}}$ is the compact SVD of the tangent vector $\tilde{\mathbf{\Delta}}$ at $\tilde{\mathbf{Y}}$ ($\tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{Y}} = \mathbf{I}$, $\tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{\Delta}} = \mathbf{0}$). For the proper complex case, the transpositions in (120) may be replaced with conjugate transpositions. Furthermore, it may be verified that for the case of geodesics provided in (120)

$$d_{\mathrm{ss}}\left(\tilde{\mathbf{Y}}(t), \tilde{\mathbf{Y}}\right) = |t|(\operatorname{tr}\, \tilde{\mathbf{\Delta}}^{\mathrm{T}}\tilde{\mathbf{\Delta}})^{\frac{1}{2}} \tag{121}$$

[see also (86) for covariance matrices and footnote 6 for the general case]. Geodesics for subspaces satisfy the differential equation $\ddot{\mathbf{Y}} + \mathbf{\Gamma}_{\mathrm{ss}}(\dot{\mathbf{Y}}, \dot{\mathbf{Y}}) = \mathbf{0}$, where the Christoffel symbols for the Grassmann manifold $G_{n,p}$ are given by the quadratic form [23]

$$\mathbf{\Gamma}_{\mathrm{ss}}(\tilde{\mathbf{\Delta}}, \tilde{\mathbf{\Delta}}) = \tilde{\mathbf{Y}}\tilde{\mathbf{\Delta}}^{\mathrm{T}}\tilde{\mathbf{\Delta}}. \tag{122}$$

The inverse exponential map is given by $\exp_{\tilde{\mathbf{Y}}}^{-1}\hat{\mathbf{Y}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$, where these matrices are computed with the compact SVDs $\mathbf{U}_1\mathbf{C}_1\mathbf{V}_1^{\mathrm{T}} := \tilde{\mathbf{Y}}^{\mathrm{T}}\hat{\mathbf{Y}}$ (singular values in ascending order) and $\mathbf{U}_2\mathbf{S}_2\mathbf{V}_2^{\mathrm{T}} := \hat{\mathbf{Y}}\mathbf{U}_1\mathbf{V}_1^{\mathrm{T}} - \tilde{\mathbf{Y}}\mathbf{V}_1\mathbf{C}_1\mathbf{V}_1^{\mathrm{T}}$ (singular values in descending order), and the formulas $\mathbf{U} = \mathbf{U}_2\mathbf{V}_2^{\mathrm{T}}\mathbf{V}_1$, $\mathbf{\Sigma} = \arccos\mathbf{C}_1$ (principal values), and $\mathbf{V} = \mathbf{V}_1$. Then, $\hat{\mathbf{Y}}\mathbf{U}_1\mathbf{V}^{\mathrm{T}} = \tilde{\mathbf{Y}}\mathbf{V}(\cos\mathbf{\Sigma})\mathbf{V}^{\mathrm{T}} + \mathbf{U}(\sin\mathbf{\Sigma})\mathbf{V}^{\mathrm{T}}$; note that $\mathbf{S}_2 = \sin\mathbf{\Sigma}$ and that $\mathbf{V}_2^{\mathrm{T}}\mathbf{V}_1 = \operatorname{diag}(\pm 1, \ldots, \pm 1)$.

C. Subspace and Covariance FIM and CRB

Theorem 1 ensures that we may extract the second-order terms of $\ell(\mathbf{Z}|\mathbf{R}_1 + \mathbf{D}, \mathbf{Y} + \mathbf{\Delta})$ in $\mathbf{D}$ and $\mathbf{\Delta}$ to compute the FIM. The Fisher information metric for the subspace plus covariance estimation problem of (1) is obtained by using the second-order terms of the Taylor series given in (62) and (63), along with (16) and the perturbation

$$(\mathbf{Y}+\mathbf{\Delta})(\mathbf{R}_1+\mathbf{D})(\mathbf{Y}+\mathbf{\Delta})^{\mathrm{T}} + \mathbf{R}_0 = \mathbf{R}_2 + \mathbf{R}_2^{(1)} + \mathbf{R}_2^{(2)} \tag{123}$$

where the first- and second-order terms are

$$\mathbf{R}_2^{(1)} = \mathbf{\Delta}\mathbf{R}_1\mathbf{Y}^{\mathrm{T}} + \mathbf{Y}\mathbf{R}_1\mathbf{\Delta}^{\mathrm{T}} + \mathbf{Y}\mathbf{D}\mathbf{Y}^{\mathrm{T}} \tag{124}$$

$$\mathbf{R}_2^{(2)} = \mathbf{\Delta}\mathbf{R}_1\mathbf{\Delta}^{\mathrm{T}} + \mathbf{\Delta}\mathbf{D}\mathbf{Y}^{\mathrm{T}} + \mathbf{Y}\mathbf{D}\mathbf{\Delta}^{\mathrm{T}}. \tag{125}$$

The resulting FIM is given by the quadratic form

$$g_{\mathrm{cov,ss}}\big((\mathbf{D}, \mathbf{\Delta}), (\mathbf{D}, \mathbf{\Delta})\big) = \frac{1}{2}K \operatorname{tr}\left((\mathbf{\Delta}\mathbf{R}_1\mathbf{Y}^{\mathrm{T}} + \mathbf{Y}\mathbf{R}_1\mathbf{\Delta}^{\mathrm{T}} + \mathbf{Y}\mathbf{D}\mathbf{Y}^{\mathrm{T}})\mathbf{R}_2^{-1}\right)^2. \tag{126}$$

Only the first-order terms of (123) appear, and (126) is consistent with the well-known case of stochastic CRBs [60]: $(\mathbf{G})_{ij} = \operatorname{tr}\left(\mathbf{R}_2^{-1}(\partial\mathbf{R}_2/\partial\theta^i)\mathbf{R}_2^{-1}(\partial\mathbf{R}_2/\partial\theta^j)\right)$ for the parameters $\theta^1, \theta^2, \ldots, \theta^n$. Applying the Woodbury formula

$$\mathbf{R}_2^{-1} = \mathbf{R}_0^{-1} - \mathbf{R}_0^{-1}\mathbf{Y}\left(\mathbf{R}_1^{-1} + \mathbf{Y}^{\mathrm{T}}\mathbf{R}_0^{-1}\mathbf{Y}\right)^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{R}_0^{-1} \tag{127}$$

to (126) and the constraint of (113) yields the simplified FIM for the covariance and subspace estimation problem:

$$g_{\mathrm{cov,ss}}\big((\mathbf{D}_1, \mathbf{\Delta}_1), (\mathbf{D}_2, \mathbf{\Delta}_2)\big) = \frac{1}{2}K \operatorname{tr}(\mathbf{I} + \mathbf{R}_1)^{-1} \times\mathbf{D}_1(\mathbf{I}+\mathbf{R}_1)^{-1}\mathbf{D}_2 + K \operatorname{tr}\, \mathbf{R}_1^2(\mathbf{I}+\mathbf{R}_1)^{-1}\mathbf{\Delta}_1^{\mathrm{T}}\mathbf{R}_0^{-1}\mathbf{\Delta}_2. \tag{128}$$

In the proper complex case, (128) may be modified by removing the factor of 1/2 that appears in the front and replacing the transpose operator with the Hermitian (conjugate transpose) operator. There are no cross terms between the covariance and subspace components of the joint FIM; therefore, there is no estimation loss on the subspace in this example.

CRBs on the covariance and subspace parameters are obtained by computing the inverse metric $g_{\mathrm{cov,ss}}^{-1}$ (as described in footnote 8). This inverse is given by the equation

$$g_{\mathrm{cov,ss}}^{-1}\big((\mathbf{U}, \mathbf{\Upsilon}), (\mathbf{U}, \mathbf{\Upsilon})\big) = g_{\mathrm{cov,ss}}\big((\mathbf{D}, \mathbf{\Delta}), (\mathbf{D}, \mathbf{\Delta})\big) \tag{129}$$

where the cotangent vectors $\mathbf{U} \in T^*_{\mathbf{R}_1}P_p$ and $\mathbf{\Upsilon} \in T^*_{\mathbf{Y}}G_{n,p}$ (colored) are defined by the equation $\mathbf{U}(\mathbf{D}') + \mathbf{\Upsilon}(\mathbf{\Delta}') = g_{\text{cov,ss}}((\mathbf{D},\mathbf{\Delta}),(\mathbf{D}',\mathbf{\Delta}'))$ for all tangent vectors $\mathbf{D}' \in T_{\mathbf{R}_1}P_p$ and $\mathbf{\Delta}' \in T_{\mathbf{Y}}G_{n,p}$ (colored). Using (76), (117), and (128) to solve for $\mathbf{D}$ and $\mathbf{\Delta}$ as functions of $\mathbf{U}$ and $\mathbf{\Upsilon}$ yields

$$\mathbf{D} = \frac{2}{K}(\mathbf{I}+\mathbf{R}_1)\mathbf{U}(\mathbf{I}+\mathbf{R}_1) \tag{130}$$

$$\mathbf{\Delta} = \frac{1}{K}\mathbf{R}_0\mathbf{\Upsilon}\mathbf{R}_1^{-2}(\mathbf{I}+\mathbf{R}_1). \tag{131}$$

Finally, these expressions for $\mathbf{D}$ and $\mathbf{\Delta}$ may be inserted into (129) to obtain the inverse Fisher information metric

$$g^{-1}_{\text{cov,ss}}\big((\mathbf{U}_1,\mathbf{\Upsilon}_1),(\mathbf{U}_2,\mathbf{\Upsilon}_2)\big) = \frac{2}{K}\operatorname{tr}(\mathbf{I}+\mathbf{R}_1)\mathbf{U}_1(\mathbf{I}+\mathbf{R}_1)\mathbf{U}_2 + \frac{1}{K}\operatorname{tr}\ \mathbf{R}_1^{-2}(\mathbf{I}+\mathbf{R}_1)\mathbf{\Upsilon}_1^{\mathrm{T}}\mathbf{R}_0\mathbf{\Upsilon}_2. \tag{132}$$

D. Computation of the Subspace CRB

The inverse Fisher information metric of (132) provides the CRB on the natural subspace distance between the true subspace and any unbiased estimate of it. Because this distance corresponds to the Frobenius norm of the elements of the matrix $\mathbf{B}$ in (115), the FIM and inverse FIM will be computed with respect to the basis of the tangent space of $G_{n,p}$ using these elements and the corresponding dual basis of the cotangent space. Using classical Cramér–Rao terminology, we will compute a lower bound on the root mean square subspace error

$$\epsilon_{\text{ss}} = E\left[d^2_{\text{ss}}(\hat{\mathbf{Y}},\mathbf{Y})\right]^{\frac{1}{2}} = E\left[\sum_{i,j}(B^{ij})^2\right]^{\frac{1}{2}} \tag{133}$$

between an estimate $\hat{\mathbf{Y}}$ of $\mathbf{Y}$, where B^{ij} are the elements of $\mathbf{B}$. The orthonormal basis vectors (whitened case) of $T_{\tilde{\mathbf{Y}}}G_{n,p}$ are

$$\{\ \tilde{\mathbf{\Delta}}_{ij} = \tilde{\mathbf{Y}}_\perp\mathbf{B}_{ij} : 1 \le i \le n-p, 1 \le j \le p\ \} \tag{134}$$

$$\tilde{\mathbf{Y}}_\perp = \text{an arbitrary } n \text{ by } (n-p) \text{ matrix such that } \tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{Y}}_\perp = \mathbf{0},\ \tilde{\mathbf{Y}}_\perp^{\mathrm{T}}\tilde{\mathbf{Y}}_\perp = \mathbf{I} \tag{135}$$

$$\mathbf{B}_{ij} = \text{an } (n-p) \text{ by } p \text{ matrix whose } ij\text{th element is unity, zeros elsewhere.} \tag{136}$$

For convenience, we will also use the orthonormal basis vectors $\mathbf{D}_{ij}$ for the tangent space of P_p defined in (74), although the invariance of the subspace accuracy to this choice of basis ensures that this choice is arbitrary.

The full FIM for the subspace and nuisance covariance parameters is written conveniently in block form as

$$\mathbf{G} = \begin{pmatrix} \mathbf{C} & \mathbf{X} \\ \mathbf{X}^{\mathrm{T}} & \mathbf{S} \end{pmatrix} \tag{137}$$

where $\mathbf{C}$ is a square matrix of order $\dim P_p$ representing the covariance block, $\mathbf{S}$ is a square matrix of order $\dim G_{n,p}$ representing the subspace block, and $\mathbf{X}$ is a $\dim P_p$ by $\dim G_{n,p}$ matrix that represents the cross terms, which vanish in this example. Using the Fisher information metric $g_{\text{cov,ss}}$ of (128), the coefficients of these blocks are

$$(\mathbf{C})_{ij;i'j'} = g_{\text{cov,ss}}\big((\mathbf{D}_{ij},\mathbf{0}),(\mathbf{D}_{i'j'},\mathbf{0})\big) \tag{138}$$

$$(\mathbf{S})_{ij;i'j'} = g_{\text{cov,ss}}\big((\mathbf{0},\mathbf{\Delta}_{ij}),(\mathbf{0},\mathbf{\Delta}_{i'j'})\big) \tag{139}$$

$$(\mathbf{X})_{ij;i'j'} = g_{\text{cov,ss}}\big((\mathbf{D}_{ij},\mathbf{0}),(\mathbf{0},\mathbf{\Delta}_{i'j'})\big). \tag{140}$$

Note that a basis for the tangent space of the product manifold $P_p \times G_{n,p}$ must be used, analogous to standard Cramér–Rao analysis with a nuisance parameter. The CRB for the subspace accuracy is given by the lower right subspace block of $\mathbf{G}^{-1}$, which is expressed using the inverse of the Schur complement of $\mathbf{C}$ as

$$\varSigma_{\mathbf{S}} = (\mathbf{S} - \mathbf{X}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{X})^{-1}. \tag{141}$$

The bound on the subspace estimation accuracy is then given by the formula

$$\epsilon^2_{\text{ss}} \ge \operatorname{tr}\ \varSigma_{\mathbf{S}} \tag{142}$$

where ϵ_{ss} is the RMSE defined in (133), and the sectional and Riemannian curvature terms $\boldsymbol{K}(\boldsymbol{b})$ and $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ have been neglected. The invariance of this result to any particular basis for the tangent space of the covariance matrices may be seen by substituting $\mathbf{C} \mapsto \mathbf{A}\mathbf{C}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{X} \mapsto \mathbf{A}\mathbf{X}$ into (141) for an arbitrary invertible matrix $\mathbf{A}$, as in standard Cramér–Rao analysis. For problems in which the cross terms $\mathbf{X}$ of (137) are nonzero, $\varSigma_{\mathbf{S}}$ in (141) quantifies the loss in measurement accuracy associated with the necessity of estimating the nuisance parameters.

Alternatively, the formula for the inverse FIM of (132) may be used to compute $\varSigma_{\mathbf{S}}$:

$$(\varSigma_{\mathbf{S}})^{ij;i'j'} = g^{-1}_{\text{cov,ss}}\left(\left(\mathbf{0},\mathbf{\Delta}^*_{ij}\right),\left(\mathbf{0},\mathbf{\Delta}^*_{i'j'}\right)\right) = K^{-1}\operatorname{tr}\ \mathbf{R}_1^{-2}(\mathbf{I}+\mathbf{R}_1)\mathbf{B}^{\mathrm{T}}_{ij}\mathbf{B}_{i'j'} \tag{143}$$

where $\mathbf{\Delta}^*_{ij} = \mathbf{R}_0^{-1/2}\tilde{\mathbf{Y}}_\perp\mathbf{B}_{ij}$ is the (colored) dual basis vector of $\mathbf{\Delta}_{ij} = \mathbf{R}_0^{1/2}\tilde{\mathbf{\Delta}}_{ij}$ in (134).

In the specific case $\mathbf{R}_1 = \text{SNR}\cdot\mathbf{I}$ (but unknown), where SNR is the signal-to-noise ratio, the blocks of the FIM in (137) with respect to the basis vectors above simplify to

$$\mathbf{S} = \frac{K\cdot\text{SNR}^2}{1+\text{SNR}}\mathbf{I}_{p(n-p)} \qquad \mathbf{C} = \frac{K\cdot\text{SNR}^2}{2(1+\text{SNR})^2}\mathbf{I}_{\frac{1}{2}p(p+1)} \qquad \mathbf{X} = \mathbf{0}. \tag{144}$$

As a result

$$\epsilon_{\text{ss}} \ge \frac{\big(p(n-p)(1+\text{SNR})\big)^{\frac{1}{2}}}{K^{\frac{1}{2}}\text{SNR}} \quad \text{(rad)} \tag{145}$$

where the sectional and Riemannian curvature terms $\boldsymbol{K}(\boldsymbol{b})$ and $\boldsymbol{R}_{\mathrm{m}}(\mathbf{C})$ have been neglected. For large SNR, this expression is well approximated by $(p(n-p))^{1/2}\times K^{-1/2}\text{SNR}^{-1/2}$ (rad).

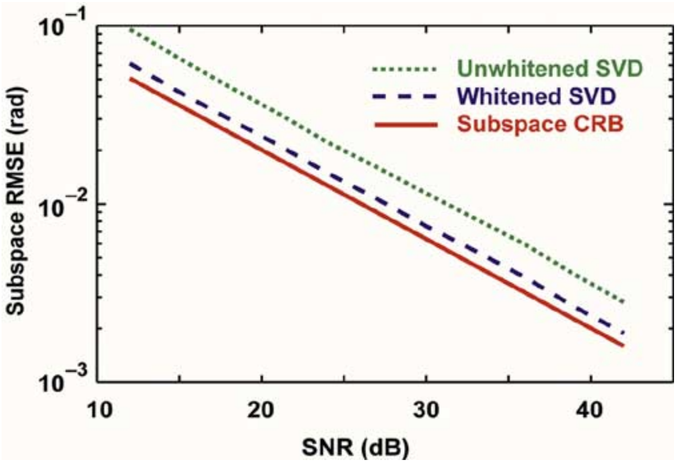

Fig. 4. RMSEs of the whitened and unwhitened SVD-based subspace estimator (Section IV-E) and the CRB of (143) versus SNR for the estimation problem of (1) on $G_{5,2}$. $\mathbf{R}_0$ and $\mathbf{R}_1$ are chosen randomly from a Wishart distribution, and $\tilde{\mathbf{Y}}$ is chosen randomly from the uniform distribution on $G_{n,p}$ [$\mathtt{orth}(\mathtt{randn}(\mathtt{n},\mathtt{p}))$ in Matlab notation]. The RMSE of the unwhitened SVD estimate is the dashed (green) curve, which exceeds the RMSE of the whitened SVD estimate (dashed blue curve) because of the bias induced by $\mathbf{R}_0$. Below these curves is the CRB in the solid (red) curve. A constant sample support of $10 = 5p$ snapshots and 1000 Monte Carlo trials are used. The RMSEs of the SVD-based subspace estimators are a small constant fraction above the Cramér–Rao lower bound over all SNRs shown.

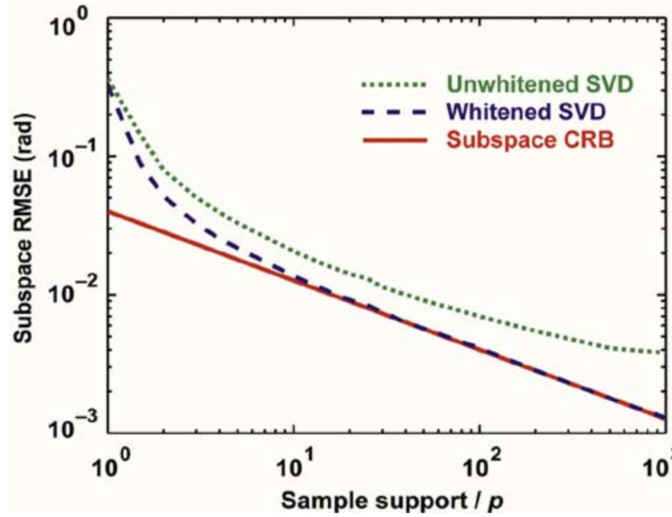

Fig. 5. RMSEs of the whitened and unwhitened subspace estimators (Section IV-E) and the CRB of (143) versus sample support (divided by p) for the estimation problem of (1) on $G_{5,2}$. The RMSE of the unwhitened SVD estimate is the dashed (green) curve, which exceeds the RMSE of the whitened SVD estimate (dashed blue curve) because of the bias induced by $\mathbf{R}_0$, especially at large sample support. Below these curves is the CRB in the solid (red) curve. A constant SNR of 21 dB and 1000 Monte Carlo trials are used. Note that, as expected, the RMSE of the whitened SVD estimate approaches the Cramér–Rao lower bound as the sample support becomes large, i.e., this maximum likelihood estimator is asymptotically efficient.

E. SVD-Based Subspace Estimation

Given an n by K data matrix $\mathbf{Z}$ whose columns are iid random vectors, the standard method [16], [28], [40] of estimating the p-dimensional principal invariant subspace of $\mathbf{Z}$ is to compute the (compact) SVD

$$\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}} := \mathbf{Z} \tag{146}$$

where $\mathbf{U}$ is an n by n orthogonal matrix, $\boldsymbol{\Sigma}$ is an n by n ordered diagonal matrix, and $\mathbf{V}$ is a K by n matrix with orthonormal columns. The p-dimensional principal invariant subspace of $\mathbf{Z}$ is taken to be the span of the first p columns of $\mathbf{U}$.

Furthermore, if an estimate is desired of the subspace represented by the matrix $\mathbf{Y}$ in (1) and the background noise covariance $\mathbf{R}_0$ is nonwhite and known, the simple SVD-based estimator using the data vectors $\mathbf{z}$ is biased by the principal invariant subspace of $\mathbf{R}_0$. In this case, a whitened SVD-based approach is used, whereby the SVD $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}} := \mathbf{R}_0^{-1/2}\mathbf{Z}$ of the whitened data matrix is computed, then $\tilde{\mathbf{Y}}$ is taken to be the first p columns of $\mathbf{U}$, and $\mathbf{Y} = \mathbf{R}_0^{1/2}\tilde{\mathbf{Y}}$. As noted, this is the maximum likelihood estimate and is therefore asymptotically efficient as $K \to \infty$ [17], [78].

F. Simulation Results

A Monte Carlo simulation was implemented to compare the subspace estimation accuracy achieved by the SVD-based estimation methods described in Section IV-E with the CRB predicted in Section IV-C. A 2-D subspace in $\mathbb{R}^5$ (chosen randomly from the uniform distribution on $G_{5,2}$) is estimated given a known 5 by 5 noise covariance $\mathbf{R}_0$ and an unknown 2 by 2 covariance matrix $\mathbf{R}_1$ (chosen randomly from a Wishart distribution, $E[\mathbf{R}_1] = \mathrm{SNR}\cdot\mathbf{I}$, where SNR is a signal-to-noise ratio that may be varied). Once $\mathbf{R}_0$, $\mathbf{R}_1$, $\mathbf{Y}$, and the number of independent snapshots K are specified, the CRB is computed from these values as described in Section IV-C. One thousand (1000) Monte Carlo trials are then performed, each of which consists of computing a normal n by K data matrix $\mathbf{Z}$ whose covariance is $\mathbf{R}_2$, estimating $\mathbf{Y}$ from the p-dimensional principal invariant subspace of $\mathbf{Z}$ and the whitened data matrix $\mathbf{R}_0^{-1/2}\mathbf{Z}$ and then computing the natural subspace distance between these estimates and $\mathbf{Y}$ (using the 2-norm of the vector of principal angles, in radians). The results comparing the accuracy of the whitened and unwhitened SVD estimators to the CRB are shown in Figs. 4 and 5 as both the SNR and the sample support K vary. As the SNR is varied, the SVD-based method achieves an accuracy that is a small constant fraction above the Cramér–Rao lower bound. Because the unwhitened SVD estimator is biased by $\mathbf{R}_0$, its RMSE error is higher than the whitened SVD estimator, especially at higher sample support. As the sample support is varied, the accuracy of the SVD-based method asymptotically approaches the lower bound, i.e., the SVD method is asymptotically efficient. We are reminded that Table II lists differential geometric objects and their more familiar counterparts in Euclidean n-space $\mathbb{R}^n$.

V. Conclusions

Covariance matrix and subspace estimation are but two examples of estimation problems on manifolds where no set of intrinsic coordinates exists. In this paper, biased and unbiased intrinsic CRBs are derived, along with several of their properties, with a view toward signal processing and related applications. Of specific applicability is an expression for the Fisher information metric that involves the second covariant differential of the log-likelihood function given an arbitrary affine connection, or,

TABLE II

ROSETTA STONE

MANIFOLD M	EUCLIDEAN n-SPACE $\mathbb{R}^n$
M manifold	$\mathbb{R}^n$ Euclidean n-space
$\boldsymbol{\theta} \in M$ point on manifold	$\mathbf{x} \in \mathbb{R}^n$ column vector
$T_{\boldsymbol{\theta}}M$ tangent space	$\mathbb{R}^n$ Euclidean n-space
$T^*_{\boldsymbol{\theta}}M$ cotangent space	$(\mathbb{R}^n)^* \cong \mathbb{R}^n$ dual Euclidean n-space
$\boldsymbol{\Omega} \in T_{\boldsymbol{\theta}}M$ tangent vector	$\mathbf{y} \in \mathbb{R}^n$ column vector
$\boldsymbol{\omega} \in T^*_{\boldsymbol{\theta}}M$ cotangent vector	$\boldsymbol{\eta} \in \mathbb{R}^{1\times n}$ 1-by-n row vector
$\boldsymbol{\theta}(t)$ curve on M, $\dot{\boldsymbol{\theta}}(t)$ tangent vector	$\mathbf{x}(t)$ curve in $\mathbb{R}^n$, $\dot{\mathbf{x}}(t)$ tangent vector
$\ell(\boldsymbol{\theta}) \in \mathbb{R}$ function on M	$\ell(\mathbf{x}) \in \mathbb{R}$ function on $\mathbb{R}^n$
$d\ell \in T^*_{\boldsymbol{\theta}}M$ differential of ℓ	$\partial\ell/\partial\mathbf{x}$ 1-by-n vector of derivatives
$d\ell(\boldsymbol{\Omega}) \in \mathbb{R}$ directional derivative of ℓ	$\partial\ell/\partial\mathbf{x}\cdot\mathbf{y}$ directional derivative of ℓ
$\boldsymbol{b}(\boldsymbol{\theta}) \in T_{\boldsymbol{\theta}}M$ vector field	$\mathbf{b}(\mathbf{x}) \in \mathbb{R}^n$ vector field
$\boldsymbol{\phi}\colon M \to P$ differentiable mapping	$\boldsymbol{\phi}(\mathbf{x})$, $\boldsymbol{\phi}\colon \mathbb{R}^n \to \mathbb{R}^p$
$\boldsymbol{\phi}_*\colon T_{\boldsymbol{\theta}}M \to T_{\boldsymbol{\phi}(\boldsymbol{\theta})}P$ push-forward	$\partial\boldsymbol{\phi}/\partial\mathbf{x}$ p-by-n Jacobian matrix of $\boldsymbol{\phi}$
$\boldsymbol{\theta} = (\theta^1, \ldots, \theta^n)$ coordinates on M	$\mathbf{x} = (x^1, \ldots, x^n)^{\mathrm{T}} \in \mathbb{R}^n$ column vector
$(\partial/\partial\theta^1)$, …, $(\partial/\partial\theta^n)$ basis of $T_{\boldsymbol{\theta}}M$	$\mathbf{e}_1$, …, $\mathbf{e}_n$ standard basis of $\mathbb{R}^n$
$d\theta^1$, …, $d\theta^n$ dual basis of $T^*_{\boldsymbol{\theta}}M$	$\mathbf{e}_1^{\mathrm{T}}$, …, $\mathbf{e}_n^{\mathrm{T}}$ standard basis of $\mathbb{R}^{1\times n}$
∇ connection/covariant differential	$\partial(\cdot)/\partial\mathbf{x}$ partial differentiation
$\nabla_{\boldsymbol{\Omega}}$ covariant derivative	$\partial(\cdot)/\partial\mathbf{x}\cdot\mathbf{y}$ directional derivative
$\nabla_{\partial/\partial\theta^i}(\partial/\partial\theta^j)=\sum_k \Gamma^k_{ij}(\partial/\partial\theta^k)$	$\partial\mathbf{e}_j/\partial\mathbf{x} = \mathbf{0} \Leftrightarrow \Gamma^k_{ij} \equiv 0$ in flat space
$\nabla\boldsymbol{b}\colon T_{\boldsymbol{\theta}}M \to T_{\boldsymbol{\theta}}M$, $\nabla\boldsymbol{b}(\boldsymbol{\Omega}) = \nabla_{\boldsymbol{\Omega}}\boldsymbol{b}$	$\partial\mathbf{b}/\partial\mathbf{x}$ n-by-n Jacobian matrix of $\mathbf{b}$
$\nabla^2\ell$ 2d covariant differential of ℓ	$\partial^2\ell/\partial\mathbf{x}^2$ n-by-n Hessian matrix of ℓ
$g\colon T_{\boldsymbol{\theta}}M \times T_{\boldsymbol{\theta}}M \to \mathbb{R}$ Riemannian metric	$\mathbf{G}$, $\mathbf{x}^{\mathrm{T}}\mathbf{G}\mathbf{x}$ quadratic form on $\mathbb{R}^n$
$g^{-1}\colon T^*_{\boldsymbol{\theta}}M \times T^*_{\boldsymbol{\theta}}M \to \mathbb{R}$ inverse metric	$\mathbf{G}^{-1}$ inverse quadratic form
$g_{ij} = g(\partial/\partial\theta^i, \partial/\partial\theta^j)$ metric coefficients	$(\mathbf{G})_{ij} = \mathbf{e}_i^{\mathrm{T}}\mathbf{G}\mathbf{e}_j$ matrix coefficients
$g^{ij} = g^{-1}(d\theta^i, d\theta^j)$ inverse coefficients	$(\mathbf{G}^{-1})^{ij} = \mathbf{e}_i^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{e}_j$ inverse coefficients
$g_{\mathrm{fim}} = E[d\ell \otimes d\ell] = -E[\nabla^2\ell]$ FIM	$\mathbf{G} = E[(\partial\ell/\partial\mathbf{x})^{\mathrm{T}}(\partial\ell/\partial\mathbf{x})] = -E[\partial^2\ell/\partial\mathbf{x}^2]$
$\mathrm{grad}_{\boldsymbol{\theta}}\,\ell \in T_{\boldsymbol{\theta}}M$ gradient vector field	$\mathrm{grad}_{\boldsymbol{\theta}}\,\ell = \mathbf{G}^{-1}(\partial\ell/\partial\mathbf{x})^{\mathrm{T}} \in \mathbb{R}^n$ gradient
$\ddot{\boldsymbol{\theta}} + \boldsymbol{\Gamma}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\theta}}) = 0$ geodesic equation	$\ddot{\mathbf{x}} = \mathbf{0}$ geodesic equation
$\boldsymbol{\Gamma}\colon T_{\boldsymbol{\theta}}M \times T_{\boldsymbol{\theta}}M \to T_{\boldsymbol{\theta}}M$	$\boldsymbol{\Gamma} \equiv \mathbf{0}$ in flat space
$\exp_{\boldsymbol{\theta}} t\boldsymbol{\Omega} \in M$ geodesic curve/exponential map	$\mathbf{x} + t\mathbf{y}$ line/vector addition
$\exp^{-1}_{\boldsymbol{\theta}} \hat{\boldsymbol{\theta}} \in T_{\boldsymbol{\theta}}M$ inverse exponential map	$\hat{\mathbf{x}} - \mathbf{x}$ vector subtraction
$\tau_t\colon T_{\boldsymbol{\theta}}M \to T_{\exp t\boldsymbol{\Omega}}M$ parallel translation	$\mathbf{y} \mapsto \mathbf{y}$ identity map is parallel
$\nabla_{\boldsymbol{\Omega}}\boldsymbol{b}\vert_{t=0} = \nabla\boldsymbol{b}(\exp t\boldsymbol{\Omega}) = \dot{\boldsymbol{b}}(0) + \boldsymbol{\Gamma}\big(\boldsymbol{b}(0), \boldsymbol{\Omega}\big)$	$(\partial\mathbf{b}/\partial\mathbf{x})\cdot\mathbf{y} = (d/dt)\vert_{t=0}\mathbf{b}(\mathbf{x} + t\mathbf{y}) = \dot{\mathbf{b}}(0)$
$\nabla_{\boldsymbol{\Omega}}\boldsymbol{b}\vert_{t=0} = \lim_{t\to 0}\big(\tau_t^{-1}\boldsymbol{b}(\exp t\boldsymbol{\Omega}) - \boldsymbol{b}(0)\big)/t$	$(\partial\mathbf{b}/\partial\mathbf{x})\cdot\mathbf{y} = \lim_{t\to 0}\big(\mathbf{b}(\mathbf{x} + t\mathbf{y}) - \mathbf{b}(0)\big)/t$
$K(\boldsymbol{X} \wedge \boldsymbol{Y}) = \langle \boldsymbol{R}(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{Y}, \boldsymbol{X}\rangle / \Vert\boldsymbol{X} \wedge \boldsymbol{Y}\Vert^2$ sectional curvature	$K \equiv 0$ in flat space
$\boldsymbol{R}(\boldsymbol{X}, \boldsymbol{Y}) = \nabla_{\boldsymbol{X}}\nabla_{\boldsymbol{Y}} - \nabla_{\boldsymbol{Y}}\nabla_{\boldsymbol{X}} - \nabla_{[\boldsymbol{X}, \boldsymbol{Y}]}$ Riemannian curvature	$\boldsymbol{R} \equiv \mathbf{0}$ in flat space

equivalently, arbitrary geodesics. These general results are then applied to the specific examples of estimating a covariance matrix in the manifold $P_n \cong \boldsymbol{Gl}(n, \mathbb{C})/\boldsymbol{U}(n)$ of positive-definite matrices and estimating a subspace in the Grassmann manifold $G_{n,p} \cong \boldsymbol{O}(n)/\big(\boldsymbol{O}(n-p) \times \boldsymbol{O}(p)\big)$ in the presence of an unknown covariance matrix in P_p. The CRB of an unbiased covariance matrix estimator using both the natural and flat (Frobenius) covariance metrics is derived and shown for the natural metric to be $(10/\log 10)n/K^{1/2}$ dB (n by n Hermitian case with sample support K). It is well known that with respect to the flat metric, the sample covariance matrix is an unbiased and efficient estimator, but this metric does not quantify the extra loss in estimation accuracy observed at low sample support. Remarkably, the sample covariance matrix is biased and not efficient with respect to the natural invariant metric on P_n, and the SCM's bias reveals the extra loss of estimation accuracy at low sample support observed in theory and practice. For this and other geometric reasons (completeness, invariance), the natural invariant metric for the covariance matrices is recommended over the flat metric for analysis of covariance matrix estimation.

The CRB of the subspace estimation problem is computed in closed form and compared with the SVD-based method of computing the principal invariant subspace of a data matrix. In the simplest case, the CRB on subspace estimation accuracy is shown to be about $(p(n-p))^{1/2}K^{-1/2}\mathrm{SNR}^{-1/2}$ rad for p-dimensional subspaces. By varying the SNR of the unknown subspace, the RMSE of the SVD-based subspace estimation method is seen to exceed the CRB by a small constant fraction. Furthermore, the SVD-based subspace estimator is confirmed to be asymptotically efficient, consistent with the fact that it is the maximum likelihood estimate. From these observations, we conclude that the principal invariant subspace can provide an excellent estimate of an unknown subspace. In addition to the examples of covariance matrices and subspaces, the theory and methods described in this paper are directly applicable to many other estimation problems on manifolds encountered in signal processing and other applications, such as computing accuracy bounds on rotation matrices, i.e., the orthogonal or unitary groups, and subspace basis vectors, i.e., the Stiefel manifold.

Finally, several intriguing open questions are suggested by the results: What is the geometric significance of the fact that only the SCM's determinant is biased? Does this fact, along with the numerical results in Section III-C, suggest improved covariance estimation techniques at low sample support? Would any such technique be preferable to the *ad hoc* but effective method of "diagonal loading"? Is the whitened SVD a biased subspace estimator? Does the geometry of the Grassmann manifold, akin to the SCM's biased determinant, have any bearing on subspace bias? Are there important applications where the curvature terms appearing in the CRB are significant? Such questions illustrate the principle that admitting a problem's geometry into its analysis not only offers a path to the problem's solution but also opens new areas of study.

ACKNOWLEDGMENT

The author benefited enormously in writing this paper from conversations with A. Hero and L. Scharf and from excellent feedback from the anonymous reviewers. In particular, Dr. Scharf's editorial and pedagogical suggestions greatly improved this paper. The author is also greatly indebted to J. Xavier, who helped him to identify an inaccuracy in the draft.

REFERENCES

[1] P.-A. Absil, R. Mahoney, and R. Sepulchre, "Riemannian geometry of Grassmann manifolds with a view on algorithmic computation," *Acta Applicandae Mathematicae*, vol. 80, no. 2, pp. 199–220, Jan. 2004.
[2] R. L. Adler, J. P. Dedieu, J. Y. Margulies, M. Martens, and M. Shub, "Newton's method on Riemannian manifolds and a geometric model for the human spine," *IMA J. Numer. Anal.*, vol. 22, pp. 359–390, 2002.
[3] S. Amari, "Differential-Geometrical Methods in Statistics," in *Lecture Notes in Statistics*. Berlin, Germany: Springer-Verlag, 1985, vol. 28.
[4] ——, "Differential geometric theory of statistics," in *Differential Geometry in Statistical Inference*, S. S. Gupta, Ed. Hayward, CA: Inst. Math. Statist., 1987, vol. 10, IMS Lecture Notes—Monograph Series.
[5] ——, *Methods of Information Geometry*. Providence, RI: Amer. Math. Soc., 2000. Originally published in Japanese as Joho kika no hoho (Tokyo, Japan: Iwanami Shoten, 1993).
[6] ——, "Information geometry on hierarchy of probability distributions," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1701–1711, Jul. 2001.
[7] W. J. Bangs, "Array Processing With Generalized Beamformers," Ph.D. dissertation, Yale Univ., New Haven, CT, 1971.
[8] T. A. Barton and S. T. Smith, "Structured covariance estimation for space-time adaptive processing," in *Proc. IEEE Conf. Acoust., Speech, Signal Process.*, Munich, Germany, 1997.
[9] R. Bhattacharya and V. Patrangenaru, "Nonparametric estimation of location and dispersion on Riemannian manifolds," *J. Statist. Planning Inference*, vol. 108, pp. 23–35, 2002.
[10] W. M. Boothby, *An Introduction to Differential Manifolds and Riemannian Geometry*, Second ed. New York: Academic, 1986.
[11] D. C. Brody and L. P. Hughston, "Statistical geometry in quantum mechanics," *Proc. R. Soc. Lond.*, vol. A 454, pp. 2445–2475, 1998.
[12] I. Chavel, *Riemannian Geometry—A Modern Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 1993.
[13] J. Cheeger and D. G. Ebin, *Comparison Theorems in Riemannian Geometry*. Amsterdam, The Netherlands: North-Holland, 1975.
[14] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. West Sussex, U.K.: Wiley, 2002.
[15] A. Cichocki and P. Georgiev, "Blind source separation algorithms with matrix constraints," *IEICE Trans. Inf. Syst.*, vol. E86-A, no. 1, pp. 522–531, Jan. 2003.
[16] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," *Proc. IEEE*, vol. 78, no. 8, pp. 1327–1343, Aug. 1990.
[17] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.
[18] P. Diaconis, "Group Representations in Probability and Statistics," in *IMS Lecture Notes—Monograph Series*, S. S. Gupta, Ed. Hayward, CA: Inst. Math. Statist., 1988, vol. 11.
[19] S. C. Douglas, "Self-stabilized gradient algorithms for blind source separation with orthogonality constraints," *IEEE Trans. Neural Networks*, vol. 11, no. 6, pp. 1490–1497, Nov. 2000.
[20] S. C. Douglas, S. Amari, and S.-Y. Kung, "On gradient adaption with unit-norm constraints," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1843–1847, Jun. 2000.
[21] A. Edelman and S. T. Smith, "On conjugate-gradient-like methods for eigen-like problems," *BIT*, vol. 36, no. 3, pp. 494–508, 1996.
[22] ——, *Linear and Nonlinear Conjugate Gradient-Related Methods*, L. Adams and J. L. Nazareth, Eds. Philadelphia, PA: SIAM, 1996.
[23] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.
[24] K.-T. Fang and Y.-T. Zhang, *Generalized Multivariate Analysis*. Beijing, China: Science Press, 1990.
[25] J. Faraut and A. Korányi, *Analysis on Symmetric Cones*. London, U.K.: Oxford Univ. Press, 1994.
[26] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Phil. Trans. Roy. Soc.*, ser. A, vol. 222, pp. 309–368, 1922, [Online] http://www.library.adelaide.edu.au/digitised/fisher/18pt1.pdf.
[27] ——, "Dispersion on a sphere," *Proc. Roy. Soc. Lond.*, ser. A, vol. 217, no. 1130, pp. 295–305, 1953.
[28] G. H. Golub and C. Van Loan, *Matrix Computations*, Third ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[29] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 1285–1301, Nov. 1990.
[30] U. Grenander, M. Miller, and A. Srivastava, "Hilbert-Schmidt lower bounds for estimators on matrix Lie groups for ATR," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 8, pp. 790–802, Aug. 1998.
[31] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*. New York: Academic, 1978.
[32] ——, *Groups and Geometric Analysis*. New York: Academic, 1984.
[33] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*. Berlin, Germany: Springer-Verlag, 1994.
[34] H. Hendriks, "A Cramér–Rao type lower bound for estimators with values on a manifold," *J. Multivariate Anal.*, vol. 38, pp. 245–261, 1991.
[35] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2002.
[36] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Stat.*, vol. 29, no. 2, pp. 295–327, 2001.
[37] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Commun. Pure Appl. Math.*, vol. 30, pp. 509–541, 1977.
[38] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*. New York: Interscience, 1969, vol. 1 and 2.
[39] J. T.-H. Lo and A. Willsky, "Estimation for rotation with one degree of freedom—Part I: Introduction and continuous time processes," *IEEE Trans. Autom. Control*, vol. AC-20, no. 1, pp. 10–21, Feb. 1975.
[40] E. Lundström and L. Eldén, "Adaptive eigenvalue computations using Newton's method on the Grassmann manifold," *SIAM Matrix Anal. Appl.*, vol. 23, no. 3, pp. 819–839, 2002.
[41] Y. Ma, J. Košecká, and S. Sastry, "Optimization criteria and geometric algorithms for motion and structure estimation," *Intl. J. Comput. Vision*, vol. 44, no. 3, pp. 219–249, 2001.
[42] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR-Sb*, vol. 1, pp. 457–483, 1967.
[43] K. V. Mardia and P. E. Jupp, *Directional Statistics*. West Sussex, U.K.: Wiley, 2000.
[44] P. McCullagh, *Tensor Methods in Statistics*. London, U.K.: Chapman and Hall, 1987.
[45] T. McWhorter and L. L. Scharf, "Cramér–Rao bounds for deterministic modal analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 5, pp. 1847–1866, May 1993.
[46] M. L. Menéndez, D. Morales, L. Pardo, and M. Salicrü, "Statistical tests based on geodesic differences," *Appl. Math. Lett.*, vol. 8, no. 1, pp. 65–69, 1995.
[47] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
[48] M. K. Murray and J. W. Rice, *Differential Geometry and Statistics*. London, U.K.: Chapman and Hall, 1993.
[49] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 7, pp. 1293–1302, Jul. 1993.
[50] A. Ohara, N. Suda, and S. Amari, "Dualistic differential geometry of positive definite matrices and its applications to related problems," *Lin. Alg. Appl.*, vol. 247, pp. 31–53, 1996.

[51] J. M. Oller, "On an intrinsic bias measure," in *Stability Problems for Stochastic Models*, V. V. Kalashnikov and V. M. Zolotarev, Eds. Berlin, Germany: Springer-Verlag, 1991, Lecture Notes in Mathematics 1546, pp. 134–158.

[52] J. M. Oller and J. M. Corcuera, "Intrinsic analysis of statistical estimation," *Ann. Stat.*, vol. 23, no. 5, pp. 1562–1581, Oct. 1995.

[53] D. Petz, "Covariance and Fisher information in quantum mechanics," *J. Phys. A: Math. Gen.*, vol. 35, pp. 929–939, 1995.

[54] K. Rahbar and J. P. Reilly, "Blind source separation algorithm for MIMO convolutive mixtures," in *Proc. Third Int. Conf. Independent Component Anal. Signal Separation*, San Diego, CA, Dec. 2001, pp. 242–247.

[55] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–89, 1945.

[56] ——, *Selected Papers of C. R. Rao*, S. Das Gupta, Ed. New York: Wiley, 1994.

[57] ——, "Differential metrics in probability spaces," in *Differential Geometry in Statistical Inference*, S. S. Gupta, Ed. Hayward, CA: Inst. Math. Statist., 1987, vol. 10, IMS Lecture Notes—Monograph Series.

[58] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 6, pp. 853–863, 1974.

[59] M. J. D. Rendas and J. M. F. Moura, "Ambiguity in radar and sonar," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 294–305, Feb. 1998.

[60] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.

[61] L. L. Scharf and T. McWhorter, "Geometry of the Cramér–Rao bound," *Signal Process.*, vol. 31, pp. 301–311, 1993.

[62] J. W. Silverstein, "The smallest eigenvalue of a large-dimensional Wishart matrix," *Ann. Prob.*, vol. 13, pp. 1364–1368, 1985.

[63] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivariate Anal.*, vol. 54, pp. 175–192, 1995.

[64] L. T. Skovgaard, "A Riemannian geometry of the multivariate normal model.," *Scand. J. Stat.*, vol. 11, pp. 211–223, 1984.

[65] S. T. Smith, "Dynamical systems that perform the singular value decomposition," *Syst. Control Lett.*, vol. 16, pp. 319–327, 1991.

[66] ——, "Subspace tracking with full rank updates," in *Proc. 31st Asilomar Conf. Signals, Syst., Comput.*, vol. 1, 1997, pp. 793–797.

[67] ——, "Adaptive radar," in *Wiley Encyclopedia of Electrical and Electronics Engineering*, J. G. Webster, Ed. New York: Wiley, 1999, vol. 1, pp. 263–289.

[68] ——, "Intrinsic Cramér–Rao bounds and subspace estimation accuracy," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, Cambridge, MA, 2000, pp. 489–493.

[69] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, Third ed. Houston, TX: Publish or Perish, 1999, vol. 1, 2.

[70] A. Srivastava, "A Bayesian approach to geometric subspace estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1390–1400, May 2002.

[71] A. Srivastava and E. Klassen, "Monte Carlo extrinsic estimators for manifold-valued parameters," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 299–308, Feb. 2002.

[72] ——, "Bayesian, geometric subspace tracking," *Advances Appl. Prob.*, to be published.

[73] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér–Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

[74] ——, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

[75] P. Stoica and B. C. Ng, "On the Cramér–Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, Jul. 1998.

[76] P. Stoica and T. Söderström, "Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies," *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1836–1847, Aug. 1991.

[77] R. J. Vaccaro and Y. Ding, "Optimal subspace-based parameter estimation," in *Proc. IEEE Conf. Acoust., Speech, Signal Process.*, vol. 4, Minneapolis, MN, USA, 1993, pp. 368–371.

[78] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968, pt. 1.

[79] J. Xavier, "Blind Identification of MIMO Channels Based on 2nd Order Statistics and Colored Inputs," Ph.D. dissertation, Instituto Superior Técnico, Lisbon, Portugal, 2002.

[80] ——, "Geodesic lower bound for parametric estimation with constraints," in *Proc. Fifth IEEE Workshop Signal Process. Adv. Wireless Commun.*, Lisbon, Portugal, Jul. 2004.

[81] R. Yang and J. O. Berger, "Estimation of a covariance matrix using the reference prior," *Ann. Stat.*, vol. 22, no. 3, pp. 1195–1211, 1994.

[82] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

Steven Thomas Smith (M'86–SM'04) was born in La Jolla, CA, in 1963. He received the B.A.Sc. degree in electrical engineering and mathematics from the University of British Columbia, Vancouver, BC, canada, in 1986 and the Ph.D. degree in applied mathematics from Harvard University, Cambridge, MA, in 1993.

From 1986 to 1988, he was a research engineer at ERIM, Ann Arbor, MI, where he developed morphological image processing algorithms. He is currently a senior member of the technical staff at MIT Lincoln Laboratory, Lexington, MA, which he joined in 1993. His research involves algorithms for adaptive signal processing, detection, and tracking to enhance radar and sonar systems performance. He has taught signal processing courses at Harvard and for the IEEE. His most recent work addresses intrinsic estimation and superresolution bounds, mean and variance CFAR, advanced tracking methods, and space-time adaptive processing algorithms.

Dr. Smith was an associate editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2000 to 2002 and received the SIAM outstanding paper award in 2001.