Hilbert Space Geometry of Quadratic Covariance Bounds

Stephen D. Howard[1], William Moran[2], Pooria Pakrooh[3] and Louis L. Scharf[3]
[1]Defence Science and Technology Group, Edinburgh, Australia
[2]School of Engineering, University of Melbourne, Melbourne, Australia
[3]Department of Mathematics, Colorado state University, Fort Collins, CO, USA

***Abstract*—In this paper, we study the geometry of quadratic covariance bounds on the estimation error covariance, in a properly defined Hilbert space of random variables. We show that a lower bound on the error covariance may be represented by the Grammian of the error score after projection onto the subspace spanned by the measurement scores. The Grammian is defined with respect to inner products in a Hilbert space of second order random variables. This geometric result holds for a large class of quadratic covariance bounds including the Barankin, Cramér-Rao, and Bhattacharyya bounds, where each bound is characterized by its corresponding measurement scores. When parameters consist of essential parameters and nuisance parameters, the Cramér-Rao covariance bound is the inverse of the Grammian of essential scores after projection onto the subspace orthogonal to the subspace spanned by the nuisance scores. In two examples, we show that for complex multivariate normal measurements with parameterized mean or covariance, there exist well-known Euclidean space geometries for the general Hilbert space geometry derived in this paper.**

I. Introduction

In [1] the authors showed that the Cramér Rao bound (CRB) [2], [3] on the variance of an unbiased estimator of a parameter θ_i in the measurement model $\mathbf{y} \sim \mathcal{N}_n(\mathbf{x}(\boldsymbol{\theta}), \sigma^2\mathbf{I})$, $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\theta} \in \mathbb{R}^p$, $p \leq n$ could be written as

$$\frac{E[(\hat{\theta}_i - \theta_i)^2]}{\sigma^2} \geq \frac{1}{\mathbf{g}_i^T \mathbf{P}_{\mathbf{G}_i}^{\perp} \mathbf{g}_i}, \tag{1}$$

where $\mathbf{g}_i = \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_i}$ characterizes the sensitivity of the mean to the i^{th} parameter, $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_p]$, $\mathbf{G}_i$ consists of all columns of $\mathbf{G}$ except $\mathbf{g}_i$, $\mathbf{P}_{\mathbf{G}_i}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathbf{G}_i}$, and $\mathbf{P}_{\mathbf{G}_i}$ is the orthogonal projection onto the subspace $\langle \mathbf{G}_i \rangle$. The denominator in (1) is the Euclidean inner product $\langle \mathbf{g}_i, \mathbf{P}_{\mathbf{G}_i}^{\perp} \mathbf{g}_i \rangle$, and the geometry is shown in Fig. 1.

This result raises the question of whether there exists a more general version of the geometry illustrated in Fig. 1. In fact, we were motivated to find a similar geometry for the case where $\boldsymbol{\theta}$ parameterizes the covariance matrix in the multivariate normal model. With this motivation, our ambition in this paper is to illuminate the geometry of the Cramér-Rao bound in a properly defined Hilbert space of random variables, and then to work two examples which show how Hilbert space inner products reduce to inner products in a properly defined Euclidean space.

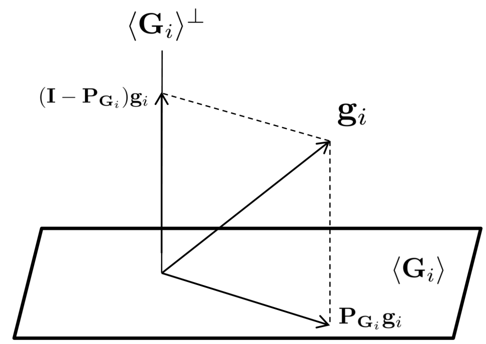

Fig. 1: Illustration of the Euclidean space geometry of the CRB in a multivariate Normal measurement model.

We start with a two channel linear estimation problem and derive the geometry of a lower bound on the Grammian of the estimation error. Then, we exploit this estimation result to discuss the geometry of quadratic covariance bounds. The quadratic covariance bounds can be derived as bounds on the minimum error covariance when linearly estimating the centered error scores from centered measurement scores. Different classes of quadratic covariance bounds are characterized by their associated measurement scores. The conceptual framework is the Hilbert space of second order random variables, but when specialized to the Cramér-Rao bound in a multivariate normal model, the Hilbert space inner products reduce to Euclidean inner products.

II. Preliminaries

Let H be a Hilbert space, with the inner product $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{C}$, and the associated norm $\| \cdot \|$. For any positive integer q denote the direct sum of q copies of H by H^q. H^q is a Hilbert space with an appropriate choice of inner product. For any ordered set of vectors $\mathbf{u} = (u_1, \ldots, u_q) \in H^q$, the Gram matrix, or Grammian $\mathbf{K}(\mathbf{u}) \in \mathbb{C}^{q \times q}$ is defined to have elements

$$\mathbf{K}(\mathbf{u})_{ij} = \langle u_i, u_j \rangle. \tag{2}$$

This work is supported in part by NSF under grant CCF-1712788, and by the Air Force Office of Scientific Research under award number FA9550-14-1-0185. Consequently the U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

Now, given another $\mathbf{v} = (v_1, \ldots, v_p) \in H^p$, we define the cross Gram matrix $\mathbf{K}(\mathbf{u};\mathbf{v})$ between the two sets as the $q \times p$ matrix with elements

$$\mathbf{K}(\mathbf{u};\mathbf{v})_{ij} = \langle u_i, v_j \rangle. \tag{3}$$

The standard inner product on H^q is $\mathrm{Tr}(\mathbf{K}(\mathbf{u};\mathbf{v}))$.

With any matrix $\mathbf{T} \in \mathbb{C}^{m \times P}$ we can associate a linear operator $\mathcal{L}_{\mathbf{T}} : H^p \to H^m$ given by

$$\mathcal{L}_{\mathbf{T}}\mathbf{v} = (\sum_{j=1}^{p} t_{1j}v_j, \sum_{j=1}^{p} t_{2j}v_j, \ldots, \sum_{j=1}^{p} t_{mj}v_j), \tag{4}$$

where $\{t_{ij}\}$ are the elements of $\mathbf{T}$.

The Gram matrix $\mathbf{K}(\mathbf{u})$ and the cross Gram matrix $\mathbf{K}(\mathbf{u};\mathbf{v})$ have the following properties:

1) $\mathbf{K}(\mathbf{u}) \succeq 0$.
2) $\mathbf{K}(\mathbf{u};\mathbf{v}) = \mathbf{K}^H(\mathbf{v};\mathbf{u})$.
3) For $c_1, c_2 \in \mathbb{C}$, $\mathbf{K}(c_1\mathbf{u}_1 + c_2\mathbf{u}_2;\mathbf{v}_1) = c_1\mathbf{K}(\mathbf{u}_1;\mathbf{v}_1) + c_2\mathbf{K}(\mathbf{u}_2;\mathbf{v}_1)$.
4) For arbitrary $\mathbf{T}_1, \mathbf{T}_2 \in \mathbb{C}^{m \times q}$ and $\mathbf{T}_3 \in \mathbb{C}^{n \times p}$, we have
   - $\mathcal{L}_{(\mathbf{T}_1+\mathbf{T}_2)}\mathbf{u} = \mathcal{L}_{\mathbf{T}_1}\mathbf{u} + \mathcal{L}_{\mathbf{T}_2}\mathbf{u}$,
   - $\mathbf{K}(\mathcal{L}_{\mathbf{T}_1}\mathbf{u}) = \mathbf{T}_1\mathbf{K}(\mathbf{u})\mathbf{T}_1^H$.
   - $\mathbf{K}(\mathcal{L}_{\mathbf{T}_1}\mathbf{u};\mathcal{L}_{\mathbf{T}_3}\mathbf{v}) = \mathbf{T}_1\mathbf{K}(\mathbf{u},\mathbf{v})\mathbf{T}_3^H$.

III. The Two-Channel Linear Estimation Experiment

Consider a two channel estimation problem where the elements of $\mathbf{u} = (u_1, \ldots, u_q)$ are to be estimated from the elements of $\mathbf{v} = (v_1, \ldots, v_p)$. For simplicity, let $\mathbf{K}_{\mathbf{uu}} = \mathbf{K}(\mathbf{u})$, $\mathbf{K}_{\mathbf{uv}} = \mathbf{K}(\mathbf{u};\mathbf{v})$, and $\mathbf{K}_{\mathbf{vv}} = \mathbf{K}(\mathbf{v})$. Consider the $(q+p) \times (q+p)$ Gram matrix $\mathbf{K}$:

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{\mathbf{uu}} & \mathbf{K}_{\mathbf{uv}} \\ \mathbf{K}_{\mathbf{uv}}^H & \mathbf{K}_{\mathbf{vv}} \end{bmatrix}. \tag{5}$$

Let $\mathbf{\Lambda} = \begin{bmatrix} \mathbf{I}_q & -\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix}$. Then $\mathbf{K}$ may be diagonalized as

$$\mathbf{\Lambda}\mathbf{K}\mathbf{\Lambda}^H = \begin{bmatrix} \mathbf{K}_{\mathbf{uu}} - \mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}\mathbf{K}_{\mathbf{uv}}^H & 0 \\ 0 & \mathbf{K}_{\mathbf{vv}} \end{bmatrix}, \tag{6}$$

where $\mathbf{I}_q$ is the identity matrix of size q. Therefore, defining $\mathbf{e} = \mathbf{u} - \mathcal{L}_{\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}}\mathbf{v}$, from (6) we have

$$\mathbf{K}(\mathbf{e}) = \mathbf{K}_{\mathbf{uu}} - \mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}\mathbf{K}_{\mathbf{uv}}^H, \tag{7}$$
$$\mathbf{K}(\mathbf{e};\mathbf{v}) = 0. \tag{8}$$

Thus, $\mathcal{L}_{\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}}\mathbf{v}$ is the best linear estimator of the elements of $\mathbf{u} = (u_1, \ldots, u_q)$ from the elements of $\mathbf{v} = (v_1, \ldots, v_p)$. That is, defining the estimation error $\zeta = \mathbf{u} - \mathcal{L}_{\mathbf{Q}}\mathbf{v}$ for any $\mathbf{Q} \in \mathbb{C}^{q \times p}$, we have

$$\begin{aligned} \mathbf{K}(\zeta) &= \mathbf{K}(\mathbf{u} - \mathcal{L}_{\mathbf{Q}}\mathbf{v}), \\ &= \mathbf{K}(\mathbf{e} + \mathcal{L}_{\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}}\mathbf{v} - \mathcal{L}_{\mathbf{Q}}\mathbf{v}) \\ &= \mathbf{K}(\mathbf{e} + \mathcal{L}_{(\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1} - \mathbf{Q})}\mathbf{v}) \\ &= \mathbf{K}(\mathbf{e}) + (\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1} - \mathbf{Q})\mathbf{K}_{\mathbf{vv}}(\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1} - \mathbf{Q})^H \\ &+ 2\mathrm{Re}\{(\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1} - \mathbf{Q})\mathbf{K}(\mathbf{v};\mathbf{e})\} \\ &\succeq \mathbf{K}(\mathbf{e}), \end{aligned} \tag{9}$$

where the last inequality comes from the facts that $\mathbf{K}(\mathbf{v};\mathbf{e}) = 0$, and $(\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1} - \mathbf{Q})\mathbf{K}_{\mathbf{vv}}(\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1} - \mathbf{Q})^H \succeq 0$. This means $\mathcal{L}_{\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}}\mathbf{v}$ is the orthogonal projection of $\mathbf{u}$ onto the subspace spanned by the elements of $\mathbf{v}$, which we write as $\mathcal{P}_{\mathbf{v}}\mathbf{u} = \mathcal{L}_{\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}}\mathbf{v}$. Furthermore,

$$\begin{aligned} \mathbf{K}(\mathcal{P}_{\mathbf{v}}\mathbf{u}) &= \mathbf{K}(\mathcal{L}_{\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}}\mathbf{v}) \\ &= \mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}\mathbf{K}_{\mathbf{vv}}(\mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1})^H \\ &= \mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}\mathbf{K}_{\mathbf{uv}}^H, \end{aligned} \tag{10}$$

where the second equality comes from property 4 in Section II, and

$$\begin{aligned} \mathbf{K}(\mathcal{P}_{\mathbf{v}}^{\perp}\mathbf{u}) &= \mathbf{K}(\mathbf{u} - \mathcal{P}_{\mathbf{v}}\mathbf{u}) \\ &= \mathbf{K}(\mathbf{e}) \\ &= \mathbf{K}_{\mathbf{uu}} - \mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}\mathbf{K}_{\mathbf{uv}}^H. \end{aligned} \tag{11}$$

Therefore, we can decompose the Grammian $\mathbf{K}_{\mathbf{uu}}$ as

$$\begin{aligned} \mathbf{K}_{\mathbf{uu}} &= \mathbf{K}(\mathcal{P}_{\mathbf{v}}\mathbf{u}) + \mathbf{K}(\mathcal{P}_{\mathbf{v}}^{\perp}\mathbf{u}) \\ &\succeq \mathbf{K}(\mathcal{P}_{\mathbf{v}}\mathbf{u}) \\ &= \mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}\mathbf{K}_{\mathbf{uv}}^H \end{aligned} \tag{12}$$

Now, consider the special case where $\mathbf{K}_{\mathbf{uv}} = [\mathbf{I}_q \; \mathbf{0}_{q \times (p-q)}]$ for some $p > q$. From (12) we have

$$\begin{aligned} \mathbf{K}_{\mathbf{uu}} &\succeq \mathbf{K}_{\mathbf{uv}}\mathbf{K}_{\mathbf{vv}}^{-1}\mathbf{K}_{\mathbf{uv}}^H \\ &= \left(\mathbf{K}_{\mathbf{vv}}^{-1}\right)_{qq}, \end{aligned} \tag{13}$$

where $\left(\mathbf{K}_{\mathbf{vv}}^{-1}\right)_{qq}$ is the northwest $q \times q$ block of $\mathbf{K}_{\mathbf{vv}}^{-1}$. Defining $\mathbf{v}_1 = (v_1, \ldots, v_q)$, $\mathbf{v}_2 = (v_{q+1}, \ldots, v_p)$, $\mathbf{K}_{\mathbf{v}_1\mathbf{v}_1} = \mathbf{K}(\mathbf{v}_1)$, $\mathbf{K}_{\mathbf{v}_2} = \mathbf{K}(\mathbf{v}_2)$, and $\mathbf{K}_{\mathbf{v}_1\mathbf{v}_2} = \mathbf{K}(\mathbf{v}_1;\mathbf{v}_2)$, we can decompose $\mathbf{K}_{\mathbf{vv}}$ as

$$\begin{aligned} \mathbf{K}_{\mathbf{vv}} &= \begin{bmatrix} \mathbf{I}_q & \mathbf{K}_{\mathbf{v}_1\mathbf{v}_2}\mathbf{K}_{\mathbf{v}_2\mathbf{v}_2}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \\ &\times \begin{bmatrix} \mathbf{K}_{\mathbf{v}_1\mathbf{v}_1} - \mathbf{K}_{\mathbf{v}_1\mathbf{v}_2}\mathbf{K}_{\mathbf{v}_2\mathbf{v}_2}^{-1}\mathbf{K}_{\mathbf{v}_1\mathbf{v}_2}^H & 0 \\ 0 & \mathbf{K}_{\mathbf{v}_2\mathbf{v}_2} \end{bmatrix} \\ &\times \begin{bmatrix} \mathbf{I}_q & 0 \\ \mathbf{K}_{\mathbf{v}_2\mathbf{v}_2}^{-1}\mathbf{K}_{\mathbf{v}_1\mathbf{v}_2}^H & \mathbf{I}_p \end{bmatrix}. \end{aligned} \tag{14}$$

Therefore, from (13) and (14), we have

$$\begin{aligned}\mathbf{K}_{\mathbf{uu}} &\succeq \left(\mathbf{K}_{\mathbf{vv}}^{-1}\right)_{qq} \\ &= \left(\mathbf{K}_{\mathbf{v}_1\mathbf{v}_1} - \mathbf{K}_{\mathbf{v}_1\mathbf{v}_2}\mathbf{K}_{\mathbf{v}_2\mathbf{v}_2}^{-1}\mathbf{K}_{\mathbf{v}_1\mathbf{v}_2}^{H}\right)^{-1} \\ &= \left(\mathbf{K}(\mathcal{P}_{\mathbf{v}_2}^{\perp}\mathbf{v}_1)\right)^{-1}, \end{aligned} \tag{15}$$

where $\mathcal{P}_{\mathbf{v}_2}^{\perp}\mathbf{v}_1 = \mathbf{v}_1 - \mathcal{P}_{\mathbf{v}_2}\mathbf{v}_1$, and $\mathcal{P}_{\mathbf{v}_2}\mathbf{v}_1 = \mathcal{L}_{\mathbf{K}_{\mathbf{v}_1\mathbf{v}_2}\mathbf{K}_{\mathbf{v}_2\mathbf{v}_2}^{1}}\mathbf{v}_2$ is the orthogonal projection of $\mathbf{v}_1 = (v_1, \ldots, v_q)$ onto the subspace spanned by the elements of $\mathbf{v}_2 = (v_{q+1}, \ldots, v_p)$. Equation (15) is our main result. The elements of $\mathbf{K}$ are Hilbert space inner products.

IV. Quadratic Covariance Bounds

Consider the complex measurement vector $\mathbf{y} \in \mathbb{C}^n$, whose distribution $f(\mathbf{y};\boldsymbol{\theta})$ is parameterized by the deterministic parameter vector $\boldsymbol{\theta} \in \mathbb{R}^p$. Let $\hat{\mathbf{g}}(\mathbf{y}) \in \mathbb{R}^q$ be an estimator of a function of parameters $\mathbf{g}(\boldsymbol{\theta}) \in \mathbb{R}^q$. Define the proper[1] estimator error as $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{y}) = \hat{\mathbf{g}}(\mathbf{y}) - \mathbf{g}(\boldsymbol{\theta})$, and the proper *centered error score* as

$$\begin{aligned}\mathbf{e}_{\boldsymbol{\theta}}(\mathbf{y}) &= \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{y}) - E_{\boldsymbol{\theta}}[\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{y})] \\ &= \hat{\mathbf{g}}(\mathbf{y}) - E_{\boldsymbol{\theta}}[\hat{\mathbf{g}}(\mathbf{y})]. \end{aligned} \tag{16}$$

Let $\boldsymbol{\sigma}_{\boldsymbol{\theta}}(\mathbf{y}) = [\sigma_1(\mathbf{y}), \ldots, \sigma_m(\mathbf{y})]^T$ be an m-dimensional vector of proper score functions, and $\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{y}) = \boldsymbol{\sigma}_{\boldsymbol{\theta}}(\mathbf{y}) - E_{\boldsymbol{\theta}}[\boldsymbol{\sigma}_{\boldsymbol{\theta}}(\mathbf{y})]$ be the *centered measurement score*, where the expected value is taken with respect to the probability density function $f(\mathbf{y};\boldsymbol{\theta})$.

Consider the Hilbert space of random variables H, with inner product defined as

$$\langle v_1, v_2 \rangle = E_{\boldsymbol{\theta}}[v_1 v_2^*] \quad \text{for} \quad v_1, v_2 \in H. \tag{17}$$

The composite covariance matrix for $[\mathbf{e}_{\boldsymbol{\theta}}^T(\mathbf{y}), \mathbf{s}_{\boldsymbol{\theta}}^T(\mathbf{y})]^T$ is

$$E_{\boldsymbol{\theta}}\left\{\begin{bmatrix}\mathbf{e}_{\boldsymbol{\theta}}(\mathbf{y}) \\ \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{y})\end{bmatrix}[\mathbf{e}_{\boldsymbol{\theta}}^H(\mathbf{y}) \; \mathbf{s}_{\boldsymbol{\theta}}^H(\mathbf{y})]\right\} = \begin{bmatrix}\mathrm{Cov}_{\boldsymbol{\theta}}\{\hat{\mathbf{g}}(\mathbf{y})\} & \mathbf{T}^H(\boldsymbol{\theta}) \\ \mathbf{T}(\boldsymbol{\theta}) & \mathbf{J}(\boldsymbol{\theta})\end{bmatrix}, \tag{18}$$

where $\mathbf{T}(\boldsymbol{\theta}) = \mathbf{K}(\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{y}); \mathbf{e}_{\boldsymbol{\theta}}(\mathbf{y}))$ is the *sensitivity matrix*, and $\mathbf{J}(\boldsymbol{\theta}) = \mathbf{K}(\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{y}))$ is the *information matrix*, both defined with respect to the inner product in (17). Based on the result in (12), we have

$$\begin{aligned}\mathrm{Cov}_{\boldsymbol{\theta}}\{\hat{g}(\mathbf{y})\} &\succeq \mathbf{T}^H(\boldsymbol{\theta})\mathbf{J}^{-1}(\boldsymbol{\theta})\mathbf{T}(\boldsymbol{\theta}) \\ &= \mathbf{K}(\mathcal{P}_{\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{y})}\mathbf{e}_{\boldsymbol{\theta}}(\mathbf{y})) \end{aligned} \tag{19}$$

where $\mathcal{P}_{\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{y})}$ is the orthogonal projection onto the subspace $\langle s_1, \ldots s_m \rangle$.[2] This formula is general. A large class of quadratic bounds on error covariance $\mathrm{Cov}_{\boldsymbol{\theta}}\{\hat{\mathbf{g}}(\mathbf{y})\}$, including the Barankin [4], Cramér-Rao [2], [3], and Bhattacharyya bounds [5] can be represented by (19), as shown in [6]. Each bound is characterized by its score function $\boldsymbol{\sigma}_{\boldsymbol{\theta}}(\mathbf{y})$. Equation (19) is derived also in [7], where it is demonstrated that score functions with zero mean which are functions of sufficient statistics for the parameters provide tighter bounds on the error covariance matrix than scores that are not zero mean, or are not functions of sufficient statistics for the parameters.

Efficiency: An efficient estimator is an estimator whose error covariance meets the lower bound in (19). That is,

$$\begin{aligned}\mathrm{Cov}_{\boldsymbol{\theta}}\{\hat{\mathbf{g}}(\mathbf{y})\} &= \mathbf{K}(\mathbf{e}_{\boldsymbol{\theta}}(\mathbf{y})) \\ &= \mathbf{K}(\mathcal{P}_{\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{y})}\mathbf{e}_{\boldsymbol{\theta}}(\mathbf{y})), \end{aligned} \tag{20}$$

which implies that the elements of $\mathbf{e}_{\boldsymbol{\theta}}(\mathbf{y})$ belong to the subspace $\langle s_1, \ldots, s_m \rangle$.

V. Fisher Score and the Cramér-Rao Bound

As a special example of the geometrical interpretation in (19), we consider the Cramér-Rao bound on the error covariance of an unbiased estimator of the parameters $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_q]^T \in \mathbb{R}^q$. The Fisher score is defined as

$$\begin{aligned}\boldsymbol{\sigma}_{\boldsymbol{\theta}}(\mathbf{y}) &= \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(\mathbf{y})\right]^H \\ &= \left[\frac{\partial}{\partial \theta_1} \log f_{\boldsymbol{\theta}}(\mathbf{y}), \ldots, \frac{\partial}{\partial \theta_p} \log f_{\boldsymbol{\theta}}(\mathbf{y})\right]^T, \end{aligned} \tag{21}$$

which has zero mean [8]. Thus, the centered Fisher score is $\mathbf{s}(\mathbf{y},\boldsymbol{\theta}) = \boldsymbol{\sigma}_{\boldsymbol{\theta}}(\mathbf{y})$, and the Fisher information matrix is $\mathbf{J}(\boldsymbol{\theta}) = E[\mathbf{s}(\mathbf{y},\boldsymbol{\theta})\mathbf{s}^H(\mathbf{y},\boldsymbol{\theta})]$. From the properties of the score function in (21), the sensitivity matrix is $\mathbf{T}(\boldsymbol{\theta}) = [\mathbf{I}_q \; \mathbf{0}_{q\times(p-q)}]$. Therefore, the general result of (19) specializes to

$$\begin{aligned}\mathrm{Cov}_{\boldsymbol{\theta}}\{\hat{\theta}(\mathbf{y})\} &\succeq \mathbf{T}^H(\boldsymbol{\theta})\mathbf{J}^{-1}(\boldsymbol{\theta})\mathbf{T}(\boldsymbol{\theta}) \\ &= \left(\mathbf{J}^{-1}(\boldsymbol{\theta})\right)_{qq}, \end{aligned} \tag{22}$$

where $\left(\mathbf{J}^{-1}(\boldsymbol{\theta})\right)_{qq}$ is the $q \times q$ northwest block of inverse of the Fisher information matrix $\mathbf{J}^{-1}(\boldsymbol{\theta})$. But from (15), this may be written as

$$\begin{aligned}\mathrm{Cov}_{\boldsymbol{\theta}}\{\hat{g}(\mathbf{y})\} &\succeq \left(\mathbf{J}^{-1}(\boldsymbol{\theta})\right)_{qq} \\ &= \left(\mathbf{K}(\mathcal{P}_{\mathbf{s}_2}^{\perp}\mathbf{s}_1)\right)^{-1}, \end{aligned} \tag{23}$$

where $\mathbf{s}_1 = (s_1, \ldots, s_q)$, $\mathbf{s}_2 = (s_{q+1}, \ldots, s_p)$ and $\mathcal{P}_{\mathbf{s}_2}^{\perp}\mathbf{s}_1 = \mathbf{s}_1 - \mathcal{P}_{\mathbf{s}_2}\mathbf{s}_1$, and $\mathcal{P}_{\mathbf{s}_2}\mathbf{s}_1$ is the orthogonal projection of the elements of $\mathbf{s}_1$ onto the subspace spanned by the elements of $\mathbf{s}_2$. The matrix $\mathbf{K}(\mathcal{P}_{\mathbf{s}_2}^{\perp}\mathbf{s}_1)$ is the Grammian of essential scores $\mathbf{s}_1$, after projection onto the subspace spanned by nuisance scores $\mathbf{s}_2$. The elements of $\mathbf{K}$ are Hilbert space inner products defined by (17). The Cramér-Rao bound on the error variance of an unbiased estimator of the parameter θ_1 is

$$\begin{aligned}\mathrm{var}_{\boldsymbol{\theta}}\{\hat{\theta}_1(\mathbf{y})\} &\geq \left(\mathbf{J}^{-1}(\boldsymbol{\theta})\right)_{11} \\ &= \frac{1}{\langle \mathcal{P}_{\mathbf{s}_2}^{\perp} s_1, \mathcal{P}_{\mathbf{s}_2}^{\perp} s_1 \rangle} \\ &= \frac{1}{\|\mathcal{P}_{\mathbf{s}_2}^{\perp} s_1\|^2}, \end{aligned} \tag{24}$$

[1] Throughout this paper, we only consider proper error and score functions. The extension of the results to the improper case is straightforward.

[2] For simplicity of notation, we show the score function as $\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{y}) = [s_1, \ldots, s_m]^T$.

where $\mathbf{s}_2 = (s_2, \ldots, s_p)$. Importantly, the denominator in (24) is a Hilbert space inner product defined by (17).

We demonstrate two examples for which there exists a Euclidean space geometry counterpart for the Hilbert space geometry of the Cramér-Rao bound.

Example 1: Complex multivariate normal measurements with parameterized mean.
Assume the measurement $\mathbf{y}$ is a proper random vector distributed as $\mathcal{CN}_n(\mathbf{x}(\boldsymbol{\theta}), \mathbf{C})$, $\mathbf{x} \in \mathbb{C}^n$, and $\boldsymbol{\theta} \in \mathbb{R}^p$. Let $\mathbf{q}_i = \mathbf{C}^{-1/2}\frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_i}$ and define $\mathbf{g}_i = (\mathbf{q}_i^T, \mathbf{q}_i^H)^T$. The $(i,j)^{th}$ element of the the Fisher information matrix is the Euclidean inner product of the $\mathbf{g}_i$ and $\mathbf{g}_j$ [1], [8]. That is

$$\begin{aligned}(\mathbf{J}(\boldsymbol{\theta}))_{ij} &= \langle \mathbf{g}_i, \mathbf{g}_j \rangle \\ &= \mathbf{g}_i^H \mathbf{g}_j. \end{aligned} \tag{25}$$

Define $\mathbf{G}_1 = [\mathbf{g}_1, \ldots, \mathbf{g}_q]$, $\mathbf{G}_2 = [\mathbf{g}_{q+1}, \ldots, \mathbf{g}_p]$. From (23), the Cramér-Rao bound on the error covariance of an unbiased estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ of $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_q]$ may be written as

$$\begin{aligned}\mathrm{Cov}_{\boldsymbol{\theta}}\{\hat{\boldsymbol{\theta}}(\mathbf{y})\} &\succeq \left(\mathbf{J}^{-1}(\boldsymbol{\theta})\right)_{qq} \\ &= \left(\mathbf{K}(\mathcal{P}_{\mathbf{G}_2}^{\perp}\mathbf{G}_1)\right)^{-1} \\ &= [\mathbf{G}_1^H(\mathbf{I} - \mathbf{P}_{\mathbf{G}_2})\mathbf{G}_1]^{-1}, \end{aligned} \tag{26}$$

where $\mathbf{P}_{\mathbf{G}_2} = \mathbf{G}_2(\mathbf{G}_2^H\mathbf{G}_2)^{-1}\mathbf{G}_2^H$ is the orthogonal projection matrix onto the subspace spanned by the columns of $\mathbf{G}_2$. In this case the Hilbert space inner products of (23) are computed as Euclidean inner products in $\mathbb{C}^{2n}$.

Example 2: Complex multivariate normal measurements with parameterized covariance.
Assume the measurement $\mathbf{y}$ is a proper random vector distributed as $\mathcal{CN}_n(\mathbf{m}, \mathbf{R}(\boldsymbol{\theta}))$, $\mathbf{m} \in \mathbb{C}^n$, and $\boldsymbol{\theta} \in \mathbb{R}^p$. Let $\mathbf{D}_i = \mathbf{R}^{-1/2}(\boldsymbol{\theta})\frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_i}\mathbf{R}^{-1/2}(\boldsymbol{\theta})$. The $(i,j)^{th}$ element of the the Fisher information matrix may be written as an inner product of $\mathbf{D}_i$ and $\mathbf{D}_j$ [8]. That is

$$\begin{aligned}(\mathbf{J}(\boldsymbol{\theta}))_{ij} &= \langle \mathbf{D}_i, \mathbf{D}_j \rangle \\ &= \mathrm{tr}(\mathbf{D}_i \mathbf{D}_j^H). \end{aligned} \tag{27}$$

Define $\mathcal{D}_1 = (\mathbf{D}_1, \ldots, \mathbf{D}_q)$, $\mathcal{D}_2 = (\mathbf{D}_{q+1}, \ldots, \mathbf{D}_p)$. Again, from (23), the Cramér-Rao bound on the error covariance of an unbiased estimator $\hat{\theta}(\mathbf{y})$ of $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_q]$ may be written as

$$\begin{aligned}\mathrm{Cov}_{\boldsymbol{\theta}}\{\hat{\boldsymbol{\theta}}(\mathbf{y})\} &\succeq \left(\mathbf{J}^{-1}(\boldsymbol{\theta})\right)_{qq} \\ &= \left(\mathbf{K}(\mathcal{P}_{\mathcal{D}_2}^{\perp}\mathcal{D}_1)\right)^{-1}, \end{aligned} \tag{28}$$

where the orthogonal projection $\mathcal{P}_{\mathcal{D}_2}^{\perp}\mathcal{D}_1$, and $\mathbf{K}(\mathcal{P}_{\mathcal{D}_2}^{\perp}\mathcal{D}_1)$ in (28) are defined with respect to the inner product in (27). Again, the Hilbert space inner products of (23) are replaced by the Euclidean inner products in $\mathbb{C}^{n\times n}$ defined in (27).

VI. Conclusion

A general class of quadratic covariance bounds on estimation error covariance may be represented as the Grammian of the error score after projection onto the space orthogonal to the subspace spanned by the measurement scores. This is the Hilbert space picture, as the Grammian is defined with respect to inner products in a Hilbert space of second order random variables. This geometric result may be applied to a large class of quadratic covariance bounds such as Barankin, Cramér-Rao, and Bhattacharyya bounds, by considering their corresponding measurement scores. In the case of Fisher score, the bound is determined by the inverse of the Grammian of essential scores after projection onto the subspace orthogonal to the subspace spanned by the nuisance scores, a result that clarifies the influence of nuisance parameters on parameter estimation.

References

[1] L. L. Scharf and L. T. McWhorter, "Geometry of the Cramér-Rao bound," *Signal Process.*, vol. 31, no. 3, pp. 301–311, Apr. 1993.

[2] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc*, vol. 37, no. 3, pp. 81–91, 1945.

[3] H. Cramér, *Mathematical Methods of Statistics (PMS-9)*, 2016, vol. 9.

[4] R. McAulay and E. Hofstetter, "Barankin bounds on parameter estimation," *IEEE Transactions on Information Theory*, vol. 17, no. 6, pp. 669–676, 1971.

[5] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation," *Sankhya: The Indian Journal of Statistics*, vol. 8, pp. 1–14, 1946.

[6] E. Weinstein and A. J. Weiss, "A general class of lower bounds in parameter estimation," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 338–342, Mar 1988.

[7] L. T. McWhorter and L. L. Scharf, "Properties of quadratic covariance bounds," in *Conf. Rec. 27th Annual Asilomar Conf. Signals, Systs., Computs., Pacific Grove, CA*, Nov. 1993, pp. 1176–1180, vol .2.

[8] P. J. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*. Cambridge University Press, 2010.