From Wirtinger to Fisher Information Inequalities on Spheres and Rotation Groups

Gregory S. Chirikjian
Department of Mechanical Engineering
Johns Hopkins University
Baltimore, MD 21218, USA

gchirik1@jhu.edu
http://rpk.lcsr.jhu.edu

***Abstract*—The concepts of Fisher Information matrix and covariance are generalized to the setting of probability densities on spheres and rotation groups, and inequalities relating these quantities are derived. Probability density functions on these spaces arise in various scenarios in the fields of structural biology, robotics, and computer vision. The approach taken is to first derive matrix generalizations of Wirtinger's inequality for tori and spheres and generalize these to rotation groups. Then new inequalities are derived that relate the covariances of probability density functions on spheres and rotation groups with their Fisher information. These inequalities are different than the Cramér-Rao bound, and can be used to estimate the rate of increase of the entropy of a diffusion process.**

***Index Terms*—Haar measure, convolution, group theory, harmonic analysis, inequalities**

I. Introduction

The Cramér-Rao bound is well known in Euclidean statistics as a way to bound the variance of an estimator with its Fisher information. However, for a variety of reasons, this bound breaks down for statistics of nonlocalized phenomena on non-Euclidean spaces. Interestingly, however, a different bound relating Fisher information and covariance is possible for spheres and rotation groups. Here this new bound is derived from Wirtinger's inequality[1], which does not have an analog for nonperiodic functions on Euclidean space.

The remainder of this introduction is broken down into three parts: A) literature review of the Cramér-Rao bound and circular statistics; B) review of the Wirtinger inequality and its generalization to spheres; C) The extension of the Wirtinger inequality to rotation groups. Later sections of this paper generalize these inequalities and apply them to generate bounds involving Fisher information and covariance which are very different than the Cramér-Rao bound.

A. Literature Review

The statistical analysis of data in Euclidean spaces is a well-developed field. A central tool in statistics and estimation in the Euclidean setting is the Cramér-Rao Bound (CRB) [2], [3], which bounds the variance of an estimator by the inverse of the Fisher information. Apparently this bound was discovered independently not only by Cramér and Rao, but also Darmois [4] and Fréchet [5]. More recently, the classical CRB has been studied from a geometric perspective [6], as well as being extended to the case of data in non-Euclidean geometric objects such as Riemannian manifolds (including circles and spheres) and Lie groups in a number of works including [7], [8]).

In this paper, an altogether different bound, which does not involve inverses, is derived for circular data based on the Wirtinger inequality. The scalar Wirtinger inequality is known for spheres as well, and is extended to both a matrix form and to the case or rotation groups in this paper. Unlike circular versions of the CRB [9], [10], [11], [12], [13], [14], [15], the inequality derived in this paper does not have an analog in Euclidean statistics.

Circular statistics has gained significant attention over the past decades, and has been summarized in various books and computer programs [16], [17], [18]. Circular data arises from bearing measurements in human-made tracking systems [19], [20], as well as in periodic signals arising from biology [21]. Other notable works on the development and/or application of circular statistics, estimation, and fusion include [22]-[25].

The concept of the circle generalizes to higher dimensions in several ways. The Cartesian product of n circles gives the n-torus (e.g., the surface of a donut when $n = 2$). The concept of bearing in the plane generalizes as a direction defined by a unit vector (or point on a sphere) in higher dimensions, and statistics on spheres has been studied extensively over more than half a century [26]. Movement around a circle generalizes as a rotation matrix in n-dimensional Euclidean space, and for $n = 3$ has applications in spacecraft attitute estimation (see [27], [28], [29], [30] for references to this vast literature). An altogether different application area that uses similar mathematical methods is the reconstruction of biomolecular structures in cryo-electron microscopy [31].

The n-sphere is an example of a Riemannian manifold, and specialized differential-geometric methods for statisitical analyisis on the sphere have been developed. The n-torus

This work was supported under ONR Grant N00014-17-1-2142, "Conformational Probabilities: A Bridge Between Innate Knowledge and Action Recognition." The author would like to thank Prof. Victor Solo for useful discussions.

[1]This inequality is given in [1] without calling it by this name, though this has now become the standard way that it is referenced.

is an Abelian Lie group, quite similar to Euclidean space, though it is compact. The rotation group, $SO(n)$, is a compact noncommutative Lie groups. The underlying manifolds for these Lie groups are also Riemannian manifolds, and could therefore be treated with the generalized tools developed for the sphere. But one can do better by using the structure of Lie groups.

B. Wirtinger's Inequality for Circles and Spheres

Wirtinger's inequality for the circle states that for a differentiable function

$$f : S^1 \to \mathbb{C} \quad \text{with} \quad \int_0^{2\pi} f(\theta)\, d\theta = 0,$$

that

$$\int_{-\pi}^{\pi} |f(\theta)|^2 d\theta \leq \int_{-\pi}^{\pi} |f'(\theta)|^2 d\theta \tag{1}$$

where $f' = df/d\theta$. The proof of (1) is straight forward: expand $f(\theta)$ as a Fourier series of the form

$$f(\theta) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \hat{f}(k)\, e^{ik\theta}$$

where

$$\hat{f}(k) = \int_{-\pi}^{\pi} f(\theta)\, e^{-ik\theta} d\theta.$$

Then take the derivative, and observe that

$$\widehat{\frac{df}{d\theta}}(k) = ik\hat{f}(k)\,.$$

Parseval's equality is written as

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\theta)|^2 d\theta \;=\; \sum_{k \in \mathbb{Z}} |\hat{f}(k)|^2\,,$$

and so

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f'(\theta)|^2 d\theta \;=\; \sum_{k \in \mathbb{Z}} k^2 |\hat{f}(k)|^2\,.$$

From the premise of the theorem, $\hat{f}(0) = 0$. Consequently, all terms in Parseval's equality for the derivative are greater than or equal to that of the original function, hence (1). Obviously, the same proof can be done in the case of an n-torus using multidimensional Fourier series to generalize (1) to

$$\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} |f(x_1, ..., x_n)|^2\, dx_1 \cdots dx_n \;\leq \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \|\nabla f\,\|^2\, dx_1 \cdots dx_n\,. \tag{2}$$

For technical reasons, throughout this paper the gradient is viewed as a row vector

$$\nabla f = \left[\frac{\partial f}{\partial x_1}\,, \cdots, \frac{\partial f}{\partial x_n} \right]\,.$$

The generalization of this to spheres gives [32]

$$\int_{S^{d-1}} |f(\mathbf{u})|^2\, d\mathbf{u} \;\leq\; \frac{1}{d-1} \int_{S^{d-1}} \|(\nabla f)(\mathbf{u})\|^2\, d\mathbf{u} \tag{3}$$

where $f : S^{d-1} \to \mathbb{C}$ is a differentiable function with

$$\int_{S^{d-1}} f(\mathbf{u})\, d\mathbf{u} = 0.$$

Here $\mathbf{u} \in S^{d-1}$ is a d-dimensional unit vector, ∇f is the gradient of f, $d\mathbf{u}$ is the usual integration measure for the sphere, and the proof follows in essentially the same way as for the circle by expanding $f(\mathbf{u})$ in hyper-spherical harmonics.

C. Wirtinger's Inequality for Rotation Groups (And Compact Lie Groups More Generally)

By the Peter-Weyl Theorem [33], matrix elements of the irreducible unitary representations (IURs) of a compact Lie group, G, form an orthonormal basis for $L^2(G)$. In the case when $G = SO(3)$, these IURs are enumerated by $l \in \mathbb{Z}_{\geq 0}$, and for any $R, A \in SO(3)$ these $(2l+1) \times (2l+1)$ IUR matrices have the fundamental properties

$$U^l(RA) = U^l(R)\, U^l(A) \;\text{ and }\; U^l(R^T) = U^l(R)^*$$

where $*$ is the Hermitian conjugate of a matrix.

For functions $f \in L^2(SO(3))$, the Fourier coefficients are computed as

$$\hat{f}^l_{mn} = \int_{SO(3)} f(A) U^l_{mn}(A^{-1})\, dA\,. \tag{4}$$

The following orthogonality relation holds

$$\int_{SO(3)} U^l_{mn}(A) \overline{U^s_{pq}(A)}\, dA = \frac{1}{2l+1} \delta_{ls} \delta_{mp} \delta_{nq} \tag{5}$$

where dA is scaled so that $\int_{SO(3)} dA = 1$. The Fourier series on $SO(3)$ has the form

$$f(A) = \sum_{l=0}^{\infty} (2l+1) \sum_{m=-l}^{l} \sum_{n=-l}^{l} \hat{f}^l_{mn} U^l_{nm}(A)\,, \tag{6}$$

which results from the completeness relation

$$\sum_{l=0}^{\infty} (2l+1) \sum_{m=-l}^{l} \sum_{n=-l}^{l} U^l_{mn}(R^{-1}) U^l_{nm}(A) = \delta(R^{-1}A)\,. \tag{7}$$

Another way to write (6) is

$$f(A) = \sum_{l=0}^{\infty} (2l+1) \text{trace} \left[\hat{f}^l\, U^l(A) \right]\,. \tag{8}$$

The Lie algebra $so(3)$ consists of skew-symmetric matrices of the form

$$X = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix} = \sum_{i=1}^{3} x_i E_i\,. \tag{9}$$

Every such matrix can be associated with a vector $\mathbf{x}$ by making the identification

$$E_i^\vee = \mathbf{e}_i \;\iff\; E_i = \hat{\mathbf{e}}_i\,.$$

This interchangeability of 3×3 skew-symmetric matrices with three-dimensional vectors via the "vee" and "hat" maps will be used extensively later.

Matrix exponentiation of skew symmetric matrices results in rotation matrices. Moreover, the logarithm of almost every $R \in SO(3)$ produces a unique $X \in so(3)$. This important fact will be used later.

The skew-symmetric matrices $\{E_i\}$ form a basis for the set of all such 3×3 skew-symmetric matrices, and the coefficients $\{x_i\}$ are all real. The $\vee$ operation extracts these coefficients from a skew symmetric matrix, X, to form a column vector $[x_1, x_2, x_3]^T \in \mathbb{R}^3$. Then $X\mathbf{y} = \mathbf{x} \times \mathbf{y}$ for any $\mathbf{y} \in \mathbb{R}^3$, where $\times$ is the usual vector cross product.

For any fixed values of x_i, it is possible to compute a directional derivative of the form

$$(\tilde{X} f)(R) \doteq \left. \frac{d}{dt} f\left(R\, e^{tX}\right) \right|_{t=0}.$$

Then evaluating at $X = E_i$ defines partial derivatives, and $(\tilde{X} f)(R) = \sum_{i=1}^{3} x_i (\tilde{E}_i f)(R)$.

The gradient of a function can be thought of as a row-vector-valued function

$$(\nabla f)(R) \doteq [(\tilde{E}_1 f)(R)\,, (\tilde{E}_2 f)(R)\,, (\tilde{E}_3 f)(R)]$$

and the Laplacian can be thought of as the scalar-valued function

$$(\nabla^2 f)(R) = (\tilde{E}_1^2 f)(R) + (\tilde{E}_2^2 f)(R) + (\tilde{E}_3^2 f)(R)$$

where $(\tilde{E}_i^2 f)(R) = (\tilde{E}_i(\tilde{E}_i f))(R)$.

It is also well-known, see e.g., [29], that the Laplacian for $SO(3)$ applied to the matrix elements of IURs gives

$$\nabla^2 U^l_{mn}(R) = -l(l+1)\, U^l_{mn}(R)\,.$$

When f is smooth enough such that ∇f and $\nabla^2 f$ both exist, then from integration by parts,

$$\int_{SO(3)} \|\nabla f(R)\|^2 dR = -\int_{SO(3)} f(R) \nabla^2 f(R)\, dR\,.$$

Combining this with the Plancherel equality

$$\int_{SO(3)} f(R)\overline{g(R)} dR = \sum_{l=0}^{\infty} (2l+1) \mathrm{trace}\left[\hat{f}^l\, (\hat{g}^l)^*\right]$$

gives

$$\int_{SO(3)} \|f(R)\|^2 dR = \sum_{l=0}^{\infty} (2l+1) \|\hat{f}^l\|^2_{HS}$$

when $f = g$, and

$$\int_{SO(3)} \|\nabla f(R)\|^2 dR = \sum_{l=0}^{\infty} (2l+1)(l+1)l \|\hat{f}^l\|^2_{HS}$$

when $g = \nabla^2 f$. (Here $\|A\|_{HS} = \sqrt{\mathrm{trace}(AA^*)}$ is the Hibert-Schmidt norm.) Consequently, when $\int_{SO(3)} f(R)\, dR = 0$, and hence $\hat{f}^0 = 0$, the inequality

$$\int_{SO(3)} |f(R)|^2\, dR \leq \int_{SO(3)} \|\nabla f(R)\|^2\, dR \qquad (10)$$

results.

The Laplacian is the *Casimir operator* for $SO(3)$. That is, it satisfies the commutation relations

$$[\nabla^2, \tilde{X}] f = [\tilde{X}, \nabla^2] f$$

for any $X \in so(3)$. The key to deriving the above expression was the use of integration by parts and the Casimir properties of the Laplacian. Since integration by parts is universal for unimodular Lie groups (including compact ones), and since Casimir operators exist for other compact Lie groups, Wirtinger's inequality can hold in more generalized contexts.

II. Fisher Information

If $\mathcal{X}$ is a measurable space with measure $dx = d\mu(x)$ evaluated at $x \in \mathcal{X}$, and if $f(x;\varphi)$ is a pdf on $\mathcal{X}$ for each φ that defines a parameterized family, then the Fisher information matrix is defined as[2]

$$F(\varphi) \doteq \int_{\mathcal{X}} \left(\nabla_\varphi \log f(x;\varphi)\right)^T \left(\nabla_\varphi \log f(x;\varphi)\right) f(x;\varphi)\, dx\,.$$

This is equivalent to

$$F(\varphi) \doteq \int_{\mathcal{X}} \frac{\left(\nabla_\varphi f(x;\varphi)\right)^T \left(\nabla_\varphi f(x;\varphi)\right)}{f(x;\varphi)}\, dx\,.$$

A. When x and φ Belong to the Same Euclidean Space or Torus

In the special case when $x, \varphi \in \mathbb{R}^n$ and

$$f(x;\varphi) = f(x - \varphi)\,,$$

then F is an $n \times n$ matrix and

$$\nabla_\varphi f(x;\varphi) = -\nabla_x f(x;\varphi)\,.$$

Moreover, due the shift invariance of integration,

$$\begin{aligned} F(\varphi) &\doteq \int_{\mathcal{X}} \frac{\left(\nabla_x f(x;\varphi)\right)^T \left(\nabla_x f(x;\varphi)\right)}{f(x;\varphi)} dx \\ &= \int_{\mathcal{X}} \frac{\left(\nabla_x f(x;0)\right)^T \left(\nabla_x f(x;0)\right)}{f(x;0)} dx \qquad (11) \\ &= F(0). \end{aligned}$$

The Cramér-Rao bound is used in Euclidean statistics to bound from above the eigenvalues of the covariance of a statistical estimator with the Fisher information of the underlying distribution. For unbiased estimators, this is written as

$$\Sigma \geq F^{-1} \qquad (12)$$

which is shorthand for $\mathbf{v}^T(\Sigma - F^{-1})\mathbf{v} \geq 0$ for all $\mathbf{v} \in \mathbb{R}^n$. It is in this sense that the matrix inequalities presented throughout this paper are expressed. A number of works have generalized this result in different ways to manifolds, including [34], [35], [36], [37].

In the case of Euclidean space, if $f(x,\varphi)$ is a Gaussian distribution with covariance Σ and mean φ, it is not difficult to show that

$$F(\varphi) = \Sigma^{-1}\,.$$

[2]As in the previous subsection, ∇_φ is considered to be a row vector.

The goal of the present work in not to write a matrix-Lie-group version of (12). Rather, inequalities of the form

$$\Sigma \leq c\,F \tag{13}$$

are derived where c is a constant that results from the structure of the space of interest. This inequality is derived in the case of compact spaces such as spheres, tori, and rotation groups via Wirtinger's inequality, which does not apply to pdfs on Euclidean space.

Moreover, if

$$S(\rho) \doteq -\int_{\mathcal{X}} \rho(x)\log\rho(x)\,dx$$

denotes the entropy of the pdf ρ, and if $f_t(x)$ denotes the solution to the diffusion equation

$$\frac{\partial f}{\partial t} = \frac{1}{2}\sum_{i,j=1}^{n} D_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j} - \sum_{k=1}^{n} d_k \frac{\partial f}{\partial x_k}$$

subject to initial conditions $f_0(x) = \delta(x)$, then the multi-dimensional extension of the de Bruijn equality states that [38]

$$\frac{d}{dt}S(\rho * f_t) = \frac{1}{2}\mathrm{tr}[DF] \tag{14}$$

where $*$ denotes convolution and $\mathrm{tr}[\cdot]$ denotes the trace of a matrix. Since Fisher information is generally difficult to compute exactly, (13) can be used to obtain lower bounds on the rate of increase of entropy. Moreover, since D is symmetric, it has an eigen-decomposition $D = Q\Lambda Q^T$ where Q is orthogonal, and since $\Lambda \leq \lambda_{max}(D)\,\mathbb{I}$, then $D \leq \lambda_{max}(D)\,\mathbb{I}$ and $\mathrm{tr}[DF] \leq \lambda_{max}(D)\,\mathrm{tr}[F]$. Consequently, the rate of increase of entropy can be bounded as

$$\frac{d}{dt}S(\rho * f_t) \leq \frac{1}{2}\lambda_{max}(D)\,\mathrm{tr}[F]\,. \tag{15}$$

B. Spheres and Rotation Groups

The pioneering work of Bingham [39], Mardia and Jupp [40], [41], [42], and Hartman and Watson [43] paved the way for the study of probability and statistics on spheres. More recently Healy, Kim, et al [44], [45] and Cui and Freeden [46] made substantial contributions. That said, there appears not to be any mention or use of Wirtinger's inequality in those works, and this affords the opportunity to present something new here.

Equipped with the appropriate concept of gradient, (12) can be defined just as easily for spheres and rotation groups as in the Euclidean case. Moreover, equipped with the appropriate concept of convolution, (14) holds as well. Convolution on unimodular Lie groups, including $SO(3)$, is defined as

$$(f_1 * f_2)(g) \doteq \int_G f_1(h) f_2(h^{-1}\circ g)\,dh\,.$$

Interestingly, with this definition of convolution, (14) holds for unimodular Lie groups in general [38], [47] with a diffusion process defined as one which satisfies the equation

$$\frac{\partial f}{\partial t} = \frac{1}{2}\sum_{i,j=1}^{n} D_{ij}\tilde{E}_i\tilde{E}_j f - \sum_{k=1}^{n} d_k \tilde{E}_k f\,.$$

subject to initial conditions $f_0(g) = \delta(g)$.

Convolution on the sphere, which is not a Lie group, can be handled as follows.

If $\rho(\mathbf{u}(\theta,\phi)) = \rho(\mathbf{u}(-\theta,\phi))$ is a pdf with mean at $\mathbf{u} = \mathbf{e}_3$, and if $R(\mathbf{a},\mathbf{b})$ is the most direct rotation that moves unit vector $\mathbf{a}$ to $\mathbf{b}$ as $\mathbf{b} = R(\mathbf{a},\mathbf{b})\,\mathbf{a}$,

$$R(\mathbf{a},\mathbf{b}) = \exp\left[\frac{\cos^{-1}(\mathbf{a}\cdot\mathbf{b})}{\|\mathbf{a}\times\mathbf{b}\|}\,\widehat{\mathbf{a}\times\mathbf{b}}\right]\,,$$

then

$$\rho(\mathbf{u};\mathbf{v}) \doteq \rho([R(\mathbf{e}_3,\mathbf{v})]^T\mathbf{u})$$

is the version of $\rho(\mathbf{u})$ with mean at $\mathbf{v}$. Then the convolution on the sphere can be defined as as

$$(\rho_1 * \rho_2)(\mathbf{u}) \doteq \int_{S^2} \rho_1(\mathbf{v})\rho_2(\mathbf{u};\mathbf{v})\,d\mathbf{v}\,.$$

Since it is possible to define a spherical gradient, and since integration of functions over the sphere is invariant under rotations,

$$\int_{S^2} f(R^T\mathbf{u})\,d\mathbf{u} = \int_{S^2} f(\mathbf{u})\,d\mathbf{u}\,,$$

and so

$$\begin{aligned} F &= \int_{S^2} \frac{\nabla_{\mathbf{u}}\rho(\mathbf{u};\mathbf{v})\nabla_{\mathbf{u}}^T\rho(\mathbf{u};\mathbf{v})}{\rho(\mathbf{u};\mathbf{v})}d\mathbf{u} \\ &= \int_{S^2} \frac{\nabla_{\mathbf{v}}\rho(\mathbf{u};\mathbf{v})\nabla_{\mathbf{v}}^T\rho(\mathbf{u};\mathbf{v})}{\rho(\mathbf{u};\mathbf{v})}d\mathbf{u}\,. \end{aligned}$$

From this, a spherical version of (14) follows using essentially the same arguements as the Euclidean case, where in this context $f_t(R)$ is a solution to a driftless diffusion equation with diffusion coefficients D_{ij}.

III. From Wirtinger to Fisher Information Inequalities

In this section, inequalities different than the Cramér-Rao bound are derived that relate Fisher information and covariance.

A. Fisher Information Inequality for the Circle

Consider the special case when $f(\theta) \doteq \theta\rho^{\frac{1}{2}}(\theta)$ with $\rho(\theta) = \rho(-\theta)$ a pdf supported in the open interval $(-a,a)$ with $a < \pi$. In this case, Wirtinger's inequality becomes

$$\int_{-\pi}^{\pi} |\theta\rho^{\frac{1}{2}}(\theta)|^2 d\theta \leq \int_{-\pi}^{\pi} \left|\rho^{\frac{1}{2}}(\theta) + \frac{1}{2}\theta\rho^{-\frac{1}{2}}(\theta)\rho'(\theta)\right|^2 d\theta\,.$$

Using the fact that $|\theta| \leq a < \pi$, and $\rho(-\pi) = \rho(\pi) = 0$, integration by parts gives

$$\int_{-\pi}^{\pi} \theta^2\rho(\theta)\,d\theta \leq \frac{a^2}{4}\int_{-\pi}^{\pi} \frac{(\rho'(\theta))^2}{\rho(\theta)}\,d\theta\,. \tag{16}$$

This bounds variance from above by Fisher information, i.e.,

$$\sigma^2 \leq \frac{a^2}{4}F\,, \tag{17}$$

In the case of the n-torus, (2) can be used to generalize (17). Letting $f(\mathbf{x}) = \mathbf{b}^T\mathbf{x}\sqrt{\rho(\mathbf{x})}$ where $\mathbf{b} \in \mathbb{R}^n$ is an arbitrary constant vector, then

$$\nabla f = \mathbf{b}^T\sqrt{\rho(\mathbf{x})} + \frac{1}{2}\mathbf{b}^T\mathbf{x}\frac{\nabla\rho(\mathbf{x})}{\sqrt{\rho(\mathbf{x})}}$$

and the result of (2) is

$$\mathbf{b}^T\Sigma\mathbf{b} \,\leq\, \mathbf{b}^T\left[\frac{1}{4}\int_{T^n}\mathbf{x}\frac{(\nabla\rho)(\nabla\rho)^T}{\rho}\mathbf{x}^T\,d\mathbf{x}\right]\mathbf{b}\,.$$

Hence,

$$\Sigma \,\leq\, \frac{1}{4}\int_{T^n}\mathbf{x}\frac{(\nabla\rho)(\nabla\rho)^T}{\rho}\mathbf{x}^T\,d\mathbf{x}\,.$$

Note that $\frac{(\nabla\rho)(\nabla\rho)^T}{\rho} = \mathrm{tr}\left[\frac{(\nabla\rho)^T(\nabla\rho)}{\rho}\right]$ is a scalar, and the matrix $\mathbf{x}\mathbf{x}^T$ has eigenvalues dominated by the inequalities $\mathbf{x}\mathbf{x}^T \leq (\mathbf{x}^T\mathbf{x})\,\mathbb{I} \leq a^2\,\mathbb{I}$ when ρ is supported on the unit ball of radius a. This gives the multidimensional version of (17) for the n-torus

$$\boxed{\Sigma \,\leq\, \frac{a^2}{4}\mathrm{tr}[F]\,\mathbb{I} \quad\Longrightarrow\quad \mathrm{tr}[\Sigma] \,\leq\, \frac{a^2 n}{4}\mathrm{tr}[F]\,.} \qquad (18)$$

As will be seen in the remainder of this paper, these inequalities are fundamental, and apply to spheres and rotation groups.

B. Fisher Information Inequality for Spheres

Given

$$f : S^{d-1} \to \mathbb{R} \;\text{ with }\; \int_{S^{d-1}} f(\mathbf{u})\,d\mathbf{u} = 0,$$

Wirtinger's inequality for the sphere in $\mathbb{R}^d$ gives [32]

$$\int_{S^{d-1}}|f(\mathbf{u})|^2\,d\mathbf{u} \leq \frac{1}{d-1}\int_{S^{d-1}}\|(\nabla f)(\mathbf{u})\|^2\,d\mathbf{u}\,. \qquad (19)$$

Here $(\nabla f)(\mathbf{u})$ is the gradient. For an m-dimensional manifold embedding in $\mathbb{R}^d$, the gradient can be defined as a row vector consisting of entries which are the projection of the vector

$$\mathrm{grad}(f) \doteq \sum_{i,j=1}^{m}\frac{\partial\mathbf{x}}{\partial q_i}g^{ij}\frac{\partial f}{\partial q_j}$$

on an orthonormal basis in the tangent plane, where $\{q_j\}$ is the set of generalized coordinates. In the present case, the position in $\mathbb{R}^d$ is $\mathbf{x} = \mathbf{u}$. Whereas in (19) it does not matter if the gradient is interpreted as a row or column vector, in the calculations that follow later, it is convenient to define it as a row vector as above.

Choosing $f(\mathbf{u}) = \mathbf{a}^T\mathbf{v}(\mathbf{u})$ where $\mathbf{v} : S^{d-1} \to \mathbb{R}^{d-1}$ gives

$$0 \leq \mathbf{a}^T\left[-\int_{S^{d-1}}\mathbf{v}\,\mathbf{v}^T\,d\mathbf{u} + \frac{1}{d-1}\int_{S^{d-1}}(\nabla\mathbf{v})(\nabla\mathbf{v})^T\,d\mathbf{u}\right]\mathbf{a}.$$

Since this is true for arbitrary $\mathbf{a}$, this can be written as the matrix (eigenvalue) inequality

$$\int_{S^{d-1}}\mathbf{v}\,\mathbf{v}^T\,d\mathbf{u} \leq \frac{1}{d-1}\int_{S^{d-1}}(\nabla\mathbf{v})(\nabla\mathbf{v})^T\,d\mathbf{u} \qquad (20)$$

This matrix version of Wirtinger's inequality can be manipulated as in the case of the circle to yield a Fisher information inequality as follows for the case of S^2.

Parameterize $\mathbf{u}$ in the usual way as

$$\mathbf{u}(\theta,\phi) = \begin{pmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{pmatrix}.$$

Then

$$d\mathbf{u} = \sin\theta\,d\phi\,d\theta\,.$$

The symmetry of $\rho(\mathbf{u})$ is written in coordinates as $\rho(\mathbf{u}(\theta,\phi)) = \rho(\mathbf{u}(\theta,\phi+\pi))$ or $\rho(\mathbf{u}(\theta,\phi)) = \rho(\mathbf{u}(-\theta,\phi))$, and the mean being at $\mathbf{u} = \mathbf{e}_3$ corresponds to $\theta = 0$. The vector $\mathbf{v}$ that describes points on the sphere relative to the origin are of the form

$$\mathbf{w}(\mathbf{u}(\theta,\phi)) = \begin{pmatrix}\theta\cos\phi\\ \theta\sin\phi\end{pmatrix}.$$

Then, by choosing

$$\mathbf{v}(\mathbf{u}) = \mathbf{w}(\mathbf{u})\,\rho^{\frac{1}{2}}(\mathbf{u})\,,$$

the left hand side of (20) defines the covariance

$$\Sigma \,=\, \int_{S^2}\mathbf{w}\mathbf{w}^T\rho\,d\mathbf{u}\,,$$

which is computed in coordinates as

$$\Sigma = \int_0^{\pi}\int_0^{2\pi}\begin{pmatrix}\cos^2\phi & \cos\phi\sin\phi\\ \cos\phi\sin\phi & \sin^2\phi\end{pmatrix}\rho(\mathbf{u}(\theta,\phi))\,\theta^2 d\mathbf{u}\,.$$

The gradient of the vector $\mathbf{v}$, is the 2×2 matrix

$$\nabla\mathbf{v} = \begin{pmatrix}\nabla v_1\\ \nabla v_2\end{pmatrix}$$

where ∇ of a scalar function is again viewed as a row vector. Specifically, the gradient of a scalar function $f(\mathbf{u})$ with $\mathbf{u} = \mathbf{u}(\theta,\phi)$ is written in coordinates as [29]:

$$\mathrm{grad}(f) = \frac{\partial f}{\partial\theta}\mathbf{e}_\theta + \frac{1}{\sin\theta}\frac{\partial f}{\partial\phi}\mathbf{e}_\phi.$$

$$\mathbf{e}_\phi = \begin{pmatrix}-\sin\phi\\ \cos\phi\\ 0\end{pmatrix}; \qquad \mathbf{e}_\theta = \begin{pmatrix}\cos\phi\cos\theta\\ \sin\phi\cos\theta\\ -\sin\theta\end{pmatrix}$$

where $\{\mathbf{e}_\theta\,,\,\mathbf{e}_\phi\}$ is an orthonormal basis for the tangent plane at $\mathbf{u}$. In this basis, the components of the gradient are expressed as the row vector

$$\nabla f = \left[\frac{\partial f}{\partial\theta}\,,\,\frac{1}{\sin\theta}\frac{\partial f}{\partial\phi}\right].$$

Applying the chain rule gives

$$\nabla\mathbf{v} = \nabla(\mathbf{w}\,\rho^{\frac{1}{2}}) = (\nabla\mathbf{w})\,\rho^{\frac{1}{2}} \,+\, \frac{1}{2}\,\mathbf{w}\,\rho^{-\frac{1}{2}}\nabla\rho\,.$$

The integrand in the right hand side of (20) then becomes

$$
\begin{aligned}
(\nabla\mathbf{v})(\nabla\mathbf{v})^T\,d\mathbf{u} \;&=\; (\nabla\mathbf{w})(\nabla\mathbf{w})^T\rho \\
&+\frac{1}{2}(\nabla\mathbf{w})(\nabla\rho)^T\mathbf{w}^T + \frac{1}{2}\mathbf{w}(\nabla\rho)(\nabla\mathbf{w})^T \\
&+\frac{1}{4}\mathbf{w}\frac{(\nabla\rho)(\nabla\rho)^T}{\rho}\mathbf{w}^T\,. \qquad (21)
\end{aligned}
$$

In coordinates

$$
\nabla\mathbf{w} = \begin{pmatrix} \cos\phi & -\frac{\theta}{\sin\theta}\sin\phi \\ \sin\phi & \frac{\theta}{\sin\theta}\cos\phi \end{pmatrix}.
$$

When ρ is supported on a region where $\theta \leq a \leq \pi/8$, then $\sin\theta \approx \theta$, and $\nabla\mathbf{w}$ is effectively a rotation matrix. Moreover, under these same conditions, the first three terms on the right side of (21) will vanish (as they did in the circle/n-torus case) after taking the trace and performing the integration over S^2.

The Fisher Information matrix for a pdf $\rho : S^2 \to \mathbb{R}_{\geq 0}$ can be defined as

$$
F \;\doteq\; \int_{S^2} \frac{(\nabla^T\rho)\,(\nabla\rho)}{\rho}\,d\mathbf{u}\,.
$$

Therefore,

$$
\int_{S^2} \frac{(\nabla\rho)(\nabla\rho)^T}{\rho}\,d\mathbf{u} = \mathrm{tr}[F]\,,
$$

and so again (18) results (with $n = 2$). If, however, a is not sufficiently small, then a different scale factor and biasing terms will result.

C. Fisher Information Inequality for the Rotation Group

The same sort of matrix Wirtinger inequality that was given for the sphere can be obtained for the rotation group as

$$
\int_{SO(3)} \mathbf{v}(R)\,\mathbf{v}^T(R)\,dR \leq \int_{SO(3)} (\nabla\mathbf{v})(R)(\nabla\mathbf{v})^T(R)\,dR \qquad (22)
$$

Let $\rho(R)$ be a pdf such that $\rho(R) = \rho(R^T)$ for all $R \in SO(3)$, i.e., ρ is a symmetric probability density function. Then necessarily, its mean will be at $R = I$. Moreover, restrict the discussion to such pdfs supported on the ball

$$
\|\log^\vee(R)\| \leq a < \pi\,. \qquad (23)
$$

Then, by choosing

$$
\mathbf{v}(R) \;\doteq\; \log^\vee(R)\,\rho^{\frac{1}{2}}(R)\,,
$$

the left hand side of (22) will become the covariance matrix

$$
\Sigma = \int_{SO(3)} \log^\vee(R)[\log^\vee(R)]^T\rho(R)\,dR\,.
$$

The right hand side of (22) is evaluated by first applying the chain rule to get the matrix

$$
(\nabla\mathbf{v})(R) \;=\; (\nabla\log^\vee)(R)\,\rho^{\frac{1}{2}}(R) + \frac{1}{2}\log^\vee(R)\,(\nabla\rho)(R)\,\rho^{-\frac{1}{2}}(R)
$$

Substituting into the right hand side of (22) then gives three terms:

$$
\begin{aligned}
\int_{SO(3)} (\nabla\mathbf{v})(R)(\nabla\mathbf{v})^T(R)\,dR =& \\
&\int_{SO(3)} (\nabla\log^\vee)(R)\,(\nabla\log^\vee)(R)\,\rho(R)\,dR \\
+\;& \frac{1}{2}\int_{SO(3)} (\nabla\log^\vee)(R)(\nabla^T\rho)(R)[\log^\vee(R)]^T\,dR \\
+\;& \frac{1}{2}\int_{SO(3)} \log^\vee(R)\,(\nabla\rho)(R)\,(\nabla\log^\vee)^T(R)\,dR \\
+\;& \frac{1}{4}\int_{SO(3)} \frac{\log^\vee(R)\,(\nabla\rho)(R)(\nabla^T\rho)(R)[\log^\vee(R)]^T}{\rho(R)}\,dR
\end{aligned}
$$

To get the desired Fisher-information-matrix inequality, it is necessary to evaluate $(\nabla\log^\vee)(R)$. This can be done by evaluating the Taylor series for the matrix logarithm, but this is not necessary because a closed-form expression exists. It is well known that the exponential map $\exp : so(3) \to SO(3)$ is related to Euler's Theorem as

$$
R = \exp(\theta\hat{\mathbf{n}}) = \mathbb{I} + \sin\theta\,\hat{\mathbf{n}} + (1-\cos\theta)\,\hat{\mathbf{n}}^2\,,
$$

where $\theta \in [0,\pi]$ is the angle of rotation around the axis $\mathbf{n} \in S^2$, with $\hat{\mathbf{n}}$ being the associated skew-symmetric matrix. Then

$$
\mathrm{tr}(R) = 1 + 2\cos\theta \quad \text{and} \quad \hat{\mathbf{n}} = \frac{R - R^T}{2\sin\theta}\,.
$$

Then, since

$$
\theta = \cos^{-1}\left[\frac{\mathrm{tr}(R)-1}{2}\right] \quad \text{and} \quad \sin(\cos^{-1}a) = \sqrt{1-a^2}\,,
$$

it follows that $\sin\theta$ can be written explicitly in terms of R as

$$
\sin\theta = \sqrt{1 - \frac{(\mathrm{tr}(R)-1)^2}{4}} = \sqrt{\frac{3}{4} - \frac{(\mathrm{tr}(R))^2}{4} + \frac{2\mathrm{tr}(R)}{4}}\,.
$$

Since $X = \theta\hat{\mathbf{n}} = \log R$, it follows that

$$
\log(R) \;=\; \frac{\cos^{-1}\left[\frac{\mathrm{tr}(R)-1}{2}\right](R - R^T)}{\sqrt{3 - (\mathrm{tr}(R))^2 + 2\mathrm{tr}(R)}}\,. \qquad (24)
$$

The exponential map, $\exp : so(3) \to SO(3)$ is surjective. However, it is not injective because there is no unique inverse for rotations by angle π around any axis. However, for all other rotations, a unique $\log$ function can be defined. Since the set on which $\log$ fails to exist is a set of measure zero, without loss of generality it is possible to exclude this set when computing integrals. Moreover, since the support of ρ is defined by the condition $\theta \leq a < \pi$, there is no issue.

The log function is odd in the sense that $\log(R^T) = -\log(R)$, as can be seen from (24). If $\rho(R)$ is even in the sense that $\rho(R) = \rho(R^T)$, then the mean of ρ is the identity element of $SO(3)$, which is the identity matrix. Moreover, if ρ is supported in a small ball around the identity, the values of the first three integrals cancel, as in the case of the circle and torus. The condition (23) means that, just like in the torus case, the eigenvalue inequality

$$
\frac{1}{4}\int_{SO(3)} \frac{\log^\vee(R)\,(\nabla\rho)(\nabla^T\rho)[\log^\vee(R)]^T}{\rho}\,dR \;\leq\; \frac{a^2}{4}\mathrm{tr}[F]\,\mathbb{I}
$$

results, and hence (18) applies to the $SO(3)$ case with $n = 3$.

IV. Conclusions

Matrix versions of Wirtinger's scalar inequality (well-known for the case of functions on circles, tori, and spheres), are derived and generalized to rotation groups. From these, new inequalities are then derived that relate the covariances of probability density functions on spheres and rotation groups with their Fisher information. These inequalities are different than the Cramér-Rao bound, and can be used to estimate the rate of increase of the entropy of a diffusion process.

References

[1] Hardy, G.I., Littlewood, J.E., Pólya, G., *Inequalities, 2nd ed.*, Cambridge Univ. Press, London and New York, 1952.

[2] Cramér, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ 1946.

[3] Rao, C.R., "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, 37:81-89, 1945.

[4] Darmois, G., "Sur les limites de la dispersion de certaines estimations," *Rev. Int. Inst. Statist.*, 13:9-15, 1945.

[5] Fréchet, M., "Sur l'extension de certaines évaluations statistiques au cas de petit échantillons," *Rev. Inst. Int. Statist.*, 11:182-205, 1943.

[6] Scharf, L.L., McWhorter, L.T., "Geometry of the Cramer-Rao bound," *Signal Processing*, 31(3):301–311, 1993.

[7] Smith, S.T., Scharf, L. McWhorter, L.T., "Intrinsic quadratic performance bounds on manifolds," *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006)*, Vol. 5, pp. V-1013-1016, May 2006.

[8] Bonnabel, S., Barrau, A., "An intrinsic Cramér-Rao bound on SO(3)for (dynamic) attitude filtering," *Proceedings of the 2015 IEEE 54th Annual Conference on Decision and Control (CDC)*, pp. 2158-2163, December 2015.

[9] Harrison, D., Kanji, G.K., Gadsden, R.J., "Analysis of Variance for Circular Data," *Journal of Applied Statistics*, 13(2):123–138, 1986.

[10] Weiss, A.J., Friedlander, B., "On The Cramér–Rao Bound for Direction Finding of Correlated Signals," *IEEE Transactions on Signal Processing*, 41(1):495–499, January 1993.

[11] Fisher, N.I., *Statistical Analysis of Circular Data*, Revised edition. Cambridge University Press, 1995.

[12] Basu, S. and Bresler, Y., "A global lower bound on parameter estimation error with periodic distortion functions," *IEEE Transactions on Information Theory*, 46(3):1145–1150, 2000.

[13] Bell, K.L., Van Trees, H.L., "Posterior Cramér–Rao Bound for Tracking Target Bearing," *13th Annual Workshop on Adaptive Sensor Array Processing (ASAP 2005)*, MIT Lincoln Lab, Lexington, MA, June 2005.

[14] Bell, K.L. and Van Trees, H.L., "Combined Cramér–Rao/Weiss–Weinstein bound for tracking target bearing," Proc. Fourth IEEE Workshop on Sensor Array and Multichannel Processing, pp. 273–277, July, 2006.

[15] Nitzan, E., Routtenberg, T. and Tabrikian, J., "Cyclic Bayesian Cramér–Rao bound for filtering in circular state space. *Proc. 18th International Conference on Information Fusion (Fusion)*, pp. 734–741, July, 2015.

[16] Jammalamadaka, S.R., Sengupta, A., *Topics in Circular Statistics*, World Scientific, Singapore, 2001.

[17] Pewsey, A., Neuhäuser, M., Ruxton, G.D., *Circular statistics in R*, Oxford University Press, 2013.

[18] Berens, P., "CircStat: a MATLAB toolbox for circular statistics," *Journal of Statistical Software*, 31(10):1–21, 2009.

[19] Kurz, G., Gilitschenski, I., Hanebeck, U.D., "Recursive nonlinear filtering for angular data based on circular distributions," *Proc. American Control Conference (ACC)*, pp. 5439–5445, June 2013.

[20] Kurz, G., Gilitschenski, I., Hanebeck, U.D., "Nonlinear measurement update for estimation of angular systems based on circular distributions," *Proc. American Control Conference (ACC)*, pp. 5694–5699, June 2014.

[21] Batschelet, E., *Circular statistics in biology*, Academic Press, New York, NY, 1981.

[22] Marcus, S.I., Kohanbash, K., "Fourier series and estimation: An application to optical phase tracking," *IEEE Transactions on Information Theory*, 24(6):775–779, Nov. 1978.

[23] Bell, K., Ephraim, Y., Van Trees, H.L., "Explicit Ziv-Zakai lower bounds for bearing estimation," *IEEE Transactions on Signal Processing*, 24:2810-2824, Nov. 1996.

[24] Azmani, M., Reboul, S., Choquel, J.-B., Benjelloun, M., "A recursive fusion filter for angular data," *Proc. of the IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 882–887, Dec. 2009.

[25] Nitzan, E., Tabrikian, J., Routtenberg, T., "Bayesian cyclic bounds for periodic parameter estimation," *Proc. of the 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 308–311, Dec. 2013.

[26] Fisher, R., "Dispersion on a Sphere," *Proceedings of the Royal Society of London A*, 217(1130):295-305, 1953.

[27] Kim, J.S., Chirikjian, G.S., "Bayesian Filtering for Orientational Distributions: A Fourier Approach," *Proc. 18th International Conference on Information Fusion (Fusion)*, pp.748-753, 2015.

[28] Wolfe, K.C., Mashner, M.J., Chirikjian, G.S., Bayesian Fusion on Lie Groups," *Journal of Algebraic Statistics* 2(1):75-97, 2011.

[29] Chirikjian, G.S., Kyatkin, A.B., *Harmonic Analysis for Engineers and Applied Scientists*, Dover, 2016.

[30] Solo, V., "Stochastic stability of an attitude estimation algorithm on SO(3)," *Proceedings of the51st Annual IEEE Conference on Decision and Control (CDC)*, pp. 1265-1270, December, 2012.

[31] Singer, A., Zhao, Z., Shkolnisky, Y. and Hadani, R., "Viewing angle classification of cryo-electron microscopy images using eigenvectors," *SIAM Journal on Imaging Sciences*, 4(2):723–759, 2011.

[32] Groemer, H., *Geometric Applications of Fourier Series and Spherical Harmonics*, Encyclopedia of Mathematics and Its Applications, Vol. 61, Cambridge University Press, 1996.

[33] Peter, F., Weyl, H., "Die Vollständigkeit der primitiven Darstellungen einer geschlossenen kontinuierlichen Gruppe," *Mathematische Annalen*, 97: 735 – 755, 1927.

[34] Xavier, J., Barroso, V., "Intrinsic variance lower bound (IVLB): an extension of the Cramer-Rao bound to Riemannian manifolds," *IEEE International Conference on Acoustics, Speech, and Signal Processing 2005 Proceedings (ICASSP 05)*, Vol. 5, pp. 10331036, March 1823, 2005

[35] Pennec, X., "Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements," *Journal of Mathematical Imaging and Vision*, 25(1): 127-154, July 2006.

[36] Hendricks, H., "A Cramer–Rao type lower bound for estimators with values in a manifold," *Journal of Multivariate Analysis*, 38:245261, 1991.

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

[38] Chirikjian, G.S., "Information–Theoretic Inequalities on Unimodular Lie Groups," *Journal of Geometric Mechanics*, 2(2):119 – 158, June 2010.

[39] Bingham, C., "An Antipodally Symmetric Distribution on the Sphere," *The Annals of Statistics*, 2(6): 1201 – 1225, 1974; Also see "Distributions on the Sphere and Projective Plane," PhD Dissertation, Yale University, 1964.

[40] Mardia, K.V., *Statistics of Directional Data*, Academic Press, New York, 1972.

[41] Jupp, P.E., Mardia, K.V., "Maximum Likelihood Estimation for the Matrix von Mises-Fisher and Bingham Distributions," *Annals of Statistics*, 7(3): 599 – 606, 1979.

[42] Khatri, C.G., Mardia, K.V., "The von Mises-Fisher Matrix Distribution in Orientation Statistics," *Journal of the Royal Statistical Society*, Series B, 39(1): 95 – 106, 1977.

[43] Hartman, P., Watson, G.S., " 'Normal' Distribution Functions on Spheres and the Modified Bessel Function," *The Annals of Probability*, 2(4): 593 – 607, 1974.

[44] Healy, D.M., Jr., Hendriks, H., Kim, P.T., "Spherical Deconvolution," *Journal of Multivariate Analysis*, 67(1): 1 – 22, 1998.

[45] Healy, D.M., Jr., Kim, P.T., "An Empirical Bayes Approach to Directional Data and Efficient Computation on the Sphere," *The Annals of Statistics*, 24(1): 232 – 254, 1996.

[46] Cui, J. Freeden, W., "Equidistribution on the sphere," *SIAM J. Sci. Comput.*, 18(2):595-609, 1997.

[47] Chirikjian, G.S., *Stochastic Models, Information Theory, and Lie Groups: Volume 2 - Analytic Methods and Modern Applications*, Birkhäuser, Boston, 2011.