Chapter 1
Information-Theoretic Matrix Inequalities and Diffusion Processes on Unimodular Lie Groups

Gregory S. Chirikjian

Abstract U
nimodular Lie groups admit natural generalizations of many of the core concepts on which classical information-theoretic inequalities are built. Specifically, they have the properties of shift-invariant integration, an associative convolution operator, well-defined diffusion processes, and concepts of Entropy, Fisher information, Gaussian distribution, and Fourier transform. Equipped with these definitions, it is shown that many inequalities from classical information theory generalize to this setting. Moreover, viewing the Fourier transform for noncommutative unimodular Lie groups as a matrix-valued function, relationships between trace inequalities, diffusion processes, and convolution are examined.

Key words: Haar measure, convolution, group theory, harmonic analysis, diffusion processes, inequalities

1.1 Introduction

This paper explores aspects of information-theoretic inequalites that naturally extend from $\mathbb{R}^n$ to an arbitrary unimodular Lie group. The exposition is mostly a condensed summary of material that can be found in [36, 35], but also presents some new inequalities. The motivations for investigating such inequalities are twofold: (1) physical information-processing agents (such as mobile robots or microscopic organisms) often have configuration spaces with Lie-group structure, and their localization in the world is therefore inextricably connected to both information theory and geometry; (2) Due to the

Johns Hopkins University, Baltimore, MD 21218, USA,
e-mail: gchirik1@jhu.edu,
WWW home page: `http://rpk.lcsr.jhu.edu`

identical form of the entropy functional in continuous information theory and in statistical mechanics, results from one field carry over to the other, and so it becomes possible to make statements about the statistical mechanical entropy of passive objects such as DNA and loops in proteins using the results of Lie-theoretic information theory.

1.1.1 Mathematical Preliminaries

A unimodular Lie group, $(G, \circ)$, is one which possesses a bi-invariant integration measure. That is, it is possible to construct a measure μ and associated volume element $dg \doteq d\mu(g)$ around each $g \in G$ such that given any function $f : G \to \mathbb{C}$ whose measure

$$\mu(f) = \int_G f(g)\, dg$$

exists, the following invariance properties will hold

$$\int_G f(g)\, dg = \int_G f(g^{-1})\, dg = \int_G f(h \circ g)\, dg = \int_G f(g \circ h)\, dg \qquad (1.1)$$

for arbitrary $h \in G$. Here, of course, g^{-1} is the inverse of $g \in G$, which is the unique element such that

$$g \circ g^{-1} = g^{-1} \circ g = e$$

with e being the identity element, which for every $g \in G$ satisifies

$$g \circ e = e \circ g = g\,.$$

This is the unique element of G with such properties.

The equalities in (1.1) are analogous to the properties of the Lebesgue integral

$$\int_{\mathbb{R}^n} f(\mathbf{x})\, d\mathbf{x} = \int_{\mathbb{R}^n} f(-\mathbf{x})\, d\mathbf{x} = \int_{\mathbb{R}^n} f(\mathbf{y} + \mathbf{x})\, d\mathbf{x} = \int_{\mathbb{R}^n} f(\mathbf{x} + \mathbf{y})\, d\mathbf{x}\,.$$

All compact Lie groups are unimodular, as are all nilpotent and semisimple Lie groups. When referring to Lie groups in this paper, the discussion is restricted to matrix Lie groups with elements that are square matrices, and group operation, $\circ$, being matrix multiplication and the identity element is the identity matrix in the case of a matrix Lie group. In this context, the set of $n \times n$ unitary matrices, $U(n)$, and all of its subgroups are compact Lie groups. An example of a noncompact semi-simple Lie group is the much-studied $SL(2, \mathbb{R})$ consisting of all 2×2 matrices with real entries and unit determinant [21, 20, 22, 23]. And an example of a nilpotent group is the Heisenberg group, $\mathcal{H}(3)$, consisting of matrices of the form

$$H(\alpha,\beta,\gamma) = \begin{pmatrix} 1 & \alpha & \beta \\ 0 & 1 & \gamma \\ 0 & 0 & 1 \end{pmatrix}$$

where $\alpha, \beta, \gamma \in \mathbb{R}$. Probability distributions, harmonic analysis, and diffusion processes on this group have been studied in detail [16, 24].

As a nontrivial example of a noncompact unimodular Lie group that arises frequently in engineering applications, and which is neither semisimple nor nilpotent, consider the Special Euclidean group, $SE(n)$, which consists of elements of the form $g = (R, \mathbf{t}) \in SO(n) \times \mathbb{R}^n$ with the semi-direct product group law

$$(R_1, \mathbf{t}_1) \circ (R_2, \mathbf{t}_2) = (R_1 R_2, R_1 \mathbf{t}_2 + \mathbf{t}_1).$$

Here $R_i \in SO(n)$, the special orthogonal group consisting of $n \times n$ rotation matrices, and the resulting semi-direct product group is denoted as

$$SE(n) = \mathbb{R}^n \rtimes SO(n)\,.$$

Building on the classic works of Miller [12, 13] and Vilenkin [19, 18, 17], the author has published extensively on harmonic analysis and diffusion processes on this group in the context of applications in robotics and polymer science [4, 6, 5]. Detailed treatment of these topics can be found in [35, 36], and a recent concise summary can be found in [1]. In order to avoid repetition, examples in the present paper are instead illustrated with $\mathcal{H}(3)$ and $SO(3)$, though the general formulation is kept abstract and general, in the spirit of [7, 9, 10, 11, 14, 15, 8, 3].

In the context of unimodular Lie groups, it then makes sense to consider probability density functions, i.e., $f : G \to \mathbb{R}$ with the properties

$$f(g) \geq 0 \ \text{ and } \int_G f(g)\, dg = 1.$$

Moreover, the concept of entropy of a pdf is simply

$$S(f) \doteq - \int_G f(g) \log f(g)\, dg\,, \tag{1.2}$$

and an entropy power is

$$N(f) \doteq \frac{1}{2\pi e} \exp\left(\frac{2}{\dim(G)} S(f)\right)\,.$$

For unimodular Lie groups, a well-defined concept of convolution of functions in $(L^1 \cap L^2)(G)$ exists:

$$(f_1 * f_2)(g) \doteq \int_G f_1(h) f_2(h^{-1} \circ g)\, dh\,. \tag{1.3}$$

This inherits the associative property from G:

$$((f_1 * f_2) * f_3)(g) = (f_1 * (f_2 * f_3))(g)\,.$$

Moreover, for broad classes of unimodular Lie groups, including all compact Lie goups and group extensions such as $SE(n)$, it is possible to define a concept of Fourier transform. This is based on the concept of an irreducible unitary representation (IUR) of G. An IUR is a unitary operator (which can be thought of as a square matrix of either finite or infinite dimension) with the proprties

$$U(g_1 \circ g_2, \lambda) = U(g_1, \lambda)\, U(g_2, \lambda) \quad \text{and} \quad U(g^{-1}, \lambda) = U^*(g, \lambda)$$

where λ can be thought of as a frequency parameter and $*$ denotes the Hermitian conjugate. The space of all Λ values is called the unitary dual of G. In the case when G is Abelian, the unitary dual is also a group. In the case of compact Lie groups, Λ is a discrete space with countable elements. In the case of some noncompact unimodular Lie groups the space Λ has been fully characterized, and its description can be a quite complicated requiring a combination of continuous and discrete parameters.

Within this context the Fourier transform is defined as

$$\hat{f}(\lambda) = \int_G f(g)\, U(g^{-1}, \lambda)\, dg\,.$$

As with classical Fourier analysis, there is a convolution theorem

$$(\widehat{f_1 * f_2})(\lambda) = \hat{f}_2(\lambda)\, \hat{f}_1(\lambda)$$

and a reconstruction formula

$$f(g) = \int_\Lambda \mathrm{tr}\left[\hat{f}(\lambda)\, U(g, \lambda)\right]\, d\lambda\,.$$

Combining the above gives

$$(f_1 * f_2)(e) = \int_\Lambda \mathrm{tr}\left[\hat{f}_2(\lambda)\, \hat{f}_1(\lambda)\right]\, d\lambda = (f_2 * f_1)(e)\,.$$

Moreover, the Plancherel equality gives

$$\int_G f_1(g)\overline{f_2(g)}\, dg = \int_\Lambda \mathrm{tr}\left[\hat{f}_1(\lambda)\, \hat{f}_2^*(\lambda)\right]\, d\lambda\,.$$

When $f_1 = f_2 = f$, this becomes the familiar form of Parseval's equality

$$\int_G |f(g)|^2 \, dg = \int_{\Lambda} \left\| \hat{f}(\lambda) \right\|_{HS}^2 \, d\lambda \,,$$

where $\|A\|_{HS}^2 = \mathrm{tr}(AA^*)$ is the familiar Hilbert-Schmidt norm.

In the context of Lie groups there are also natural generalizations of the concept of partial derivatives. Namely, if $X \in \mathcal{G}$ (the Lie algebra corresponding to the Lie group G), then a (left-invariant) directional derivative of $f(g)$ is computed as

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

If $\{E_i\}$ is a basis for $\mathcal{G}$, then $(\tilde{E}_i f)(g)$ can be thought of as partial derivatives, and the derivative in the direction $X = \sum_i x_i E_i$ can be written as

$$(\tilde{X} f)(g) = \sum_i x_i (\tilde{E}_i f)(g) \,.$$

Making a choice $\{E_i\}$ and defining an inner product by imposing orthonormality conditions $(E_i, E_j) = \delta_{ij}$ in effect fixes a metric for $\mathcal{G}$, which can be transported by left or right action to define a metric on G.

Operational properties of the Fourier transform include

$$\widehat{(\tilde{X} f)}(\lambda) = u(X, \lambda)\, \hat{f}(\lambda)$$

where

$$u(X, \lambda) \doteq \frac{d}{dt} U(e^{tX}, \lambda)\Big|_{t=0} \,.$$

This matrix function is linear in X, and so

$$u\left(\sum_i x_i E_i, \lambda\right) = \sum_i x_i u(E_i, \lambda) \,.$$

The concept of a Fisher information matix with elements

$$F_{ij}(f) \doteq \int_G \frac{(\tilde{E}_i f)(g)(\tilde{E}_j f)(g)}{f(g)} \, dg$$

and a diffusion process on G can be described with an equation of the form[1]

$$\frac{\partial f}{\partial t} = -\sum_{i=1}^{dim(G)} h_i \tilde{E}_i f + \frac{1}{2} \sum_{i,j=1}^{dim(G)} D_{ij} \tilde{E}_i \tilde{E}_j f \,. \tag{1.4}$$

[1] The diffusion coefficients are D_{ij} and the drift coefficients are h_i. When $h_i = 0$ for all $i \in \{1, ..., dim(G)\}$ the diffusion process is called driftless.

With all of this in mind, it becomes possible to explore generalizations of inequalities from classical information theory. Two major hindrances to trivially extending these inequalities to the noncommutative case are: (1) for general functions on G,

$$(f_1 * f_2)(g) \neq (f_2 * f_1)(g)\,;$$

and, (2) unlike in the case of $\mathbb{R}^n$ where the Gaussian distribution is simultaneously (a) the maximal entropy distribution subject to covariance constraints, (b) is closed under convolution and conditioning, and (c) solves the Euclidean-space version of (1.4), these attributes cannot in general be simultaneously satisfied for more general Lie groups. For example, the Entropy-Power Inequality (EPI) is not even true for the circle group $\mathbb{R}/\mathbb{Z} \cong SO(2)$.

This paper summarizes what is already known, introduces some new results, and poses some questions for future exploration.

1.1.2 Structure of The Paper

Section 1.2 explains how to compute the integration measure for unimodular Lie groups. Section 1.3 reviews the concept of functions of positive type on unimodular Lie groups, and the resulting properties of Fourier matrices for such functions. Section 1.4 explains why trace inequalities are significant in the harmonic analysis of diffusion processes on noncommutative unimodular Lie groups, and reviews some of the most well-known trace inequalities and various conjectures put forth in the literature. Section 1.5 reviews how Fisher information arises in quantifying entropy increases under diffusion and reviews a generalization the de Bruijn identity, which is shown to hold for unimodular Lie groups in general. This too involves trace inequalities. Section 1.6 reviews definitions of mean of covariance of probability densities on unimodular Lie groups, and how the propagate under convolution. Section 1.7 illustrates the theory with specific examples ($SO(3)$ as an example of compact Lie groups and the Heisenberg group $\mathcal{H}(3)$ as an example of a noncommutative noncompact unimodular Lie group.)

1.2 Explicit Computation of the Bi-Invariant Integration Measure

This section explains how to compute the bi-invariant integration measure for a unimodular Lie group, summarizing the discussion in [36, 35].

To begin, an inner product $(\cdot,\cdot)$ between arbitrary elements of the Lie algebra, $Y = \sum_i y_i E_i$ and $Z = \sum_j z_j E_j$, can be defined such that

$$(Y,Z) \doteq \sum_{i=1}^{n} y_i z_i \quad \text{where} \quad (E_i, E_j) = \delta_{ij}. \tag{1.5}$$

The basis $\{E_i\}$ is then orthonormal with respect to this inner product. The definition of the inner product together with the constraint of orthonormality of a particular choice of basis $\{E_i\}$ in (1.5) defines a metric tensor for the Lie group.

Let $\mathbf{q} = [q_1, ..., q_n]^T$ be a column vector of local coordinates. Then $g(t) = \tilde{g}(\mathbf{q}(t))$ is a curve in G where $\tilde{g} : \mathbb{R}^n \to G$ is the local parametrization of the Lie group G. Henceforth the tilde will be dropped since it will be clear from the argument whether the function $g(t)$ or $g(\mathbf{q})$ is being referred to. The right-Jacobian matrix[2] for an n-dimensional Lie group parameterized with local coordinates $q_1, ..., q_n$ is the matrix $J_r(\mathbf{q})$ that relates rates of change $\dot{\mathbf{q}}$ to $g^{-1}\dot{g}$, and likewise for $J_l(\mathbf{q})$ and $\dot{g}g^{-1}$, where a dot denotes d/dt. Specifically,

$$\dot{g}g^{-1} = \sum_j \omega_j^l E_j \quad \text{and} \quad \boldsymbol{\omega}^l = J_l(\mathbf{q})\dot{\mathbf{q}}$$

and

$$g^{-1}\dot{g} = \sum_j \omega_j^r E_j \quad \text{and} \quad \boldsymbol{\omega}^r = J_r(\mathbf{q})\dot{\mathbf{q}}.$$

In other words,

$$(\dot{g}g^{-1}, E_k) = \left(\sum_j \omega_j^l E_j, E_k \right) = \sum_j \omega_j^l \,(E_j, E_k) = \sum_j \omega_j^l \delta_{jk} = \omega_k^l.$$

The scalars ω_k^l can be stacked in an array to form the column vector $\boldsymbol{\omega}^l = [\omega_1^l, \omega_2^l, ..., \omega_n^l]^T$. Analogous calculations follow for the "r" case. This whole process is abbreviated with the "$\vee$" operation as

$$\left(\dot{g}g^{-1}\right)^{\vee} = \boldsymbol{\omega}^l \quad \text{and} \quad \left(g^{-1}\dot{g}\right)^{\vee} = \boldsymbol{\omega}^r. \tag{1.6}$$

Given an orthogonal basis $E_1, ..., E_n$ for the Lie algebra, projecting the left and right tangent operators onto this basis yields elements of the right- and left-Jacobian matrices:[3]

[2] Here 'right' and 'left' respectively refer to differentiation appearing on the right or left side in calculations. As such a 'right' quantity denoted with a subscript r is left invariant, and a 'left' quantity denoted with a subscript l is right invariant.

[3] The 'l' and 'r' convention used here for Jacobians and for vector fields is opposite that used in the mathematics literature. The reason for the choice made here is to emphasize the location of the "the most informative part" of the expression. In Jacobians, this is the location of the partial derivatives. In vector fields this is where the components defining the field appear.

$$(J_r)_{ij} = \left(g^{-1}\frac{\partial g}{\partial q_j}, E_i\right) \quad \text{and} \quad (J_l)_{ij} = \left(\frac{\partial g}{\partial q_j}g^{-1}, E_i\right). \tag{1.7}$$

In terms of the $\vee$ operation this is written as

$$\left(g^{-1}\frac{\partial g}{\partial q_j}\right)^{\vee} = J_r(\mathbf{q})\,\mathbf{e}_j \quad \text{and} \quad \left(\frac{\partial g}{\partial q_j}g^{-1}\right)^{\vee} = J_l(\mathbf{q})\,\mathbf{e}_j.$$

As another abuse of notation, the distinction between $J(\mathbf{q})$ and $J(g(\mathbf{q}))$ can be blurred in both the left and right cases. Again, it is clear which is being referred to from the argument of these matrix-valued functions.

Note that $J_r(h \circ g) = J_r(g)$ and $J_l(g \circ h) = J_l(g)$. For unimodular Lie groups,

$$|\det(J_r)(\mathbf{q})| \;=\; |\det(J_l)(\mathbf{q})| \quad \text{and} \quad dg = |\det(J_{r,l})(\mathbf{q})|\,d\mathbf{q}\,. \tag{1.8}$$

This dg has the bi-invariance property, and is called the *Haar measure*. Examples of how this looks in different coordinates are given for $\mathcal{H}(3)$ and $SO(3)$ in Section 1.7. In the compact case, it is always possible to find a constant c to normalize as $d'g \doteq c \cdot dg$ such that $\int_G d'g = 1$.

1.3 Functions of Positive Type

In harmonic analysis, a function $\varphi : G \to \mathbb{C}$ is called a *function of positive type* if for every $c_i \in \mathbb{C}$ and every $g_i, g_j \in G$ and any $n \in \mathbb{Z}_{>0}$ the inequality

$$\sum_{i,j=1}^{n} c_i\,\overline{c_j}\,\varphi(g_i \circ g_j^{-1}) \;\geq\; 0\,.$$

In some texts, such functions are also called *positive definite*, whereas in others that term is used only when the inequality above excludes equality except when all values of c_i are zero. Here a function of positive type will be taken to be one for which the matrix $M = [m_{ij}]$ with entries $m_{ij} \doteq \varphi(g_i \circ g_j^{-1})$ is Hermitian positive semi-definite (which can be shown to be equivalent to the above expression), and a positive definite function is one for which M is positive definite.

Some well-known properties of functions of positive type include [7, 9, 11]:

$$\begin{aligned} \varphi(e) &= \overline{\varphi(e)} \geq 0 \\ |\varphi(g)| &\leq \varphi(e) \\ \varphi(g^{-1}) &= \overline{\varphi(g)}\,. \end{aligned}$$

Moreover, if φ_1 and φ_2 are two such functions, then so are $\overline{\varphi_i}$, $\varphi_1 \cdot \varphi_2$, as are linear combinations of the form $a_1\varphi_1 + a_2\varphi_2$ where $a_i \in \mathbb{R}_{>0}$.

Clearly, if φ is a function constructed as

$$\varphi(g;\lambda) \doteq \mathrm{tr}\,[A^* U(g,\lambda)A]$$

when A is positive definite, then

$$\sum_{i,j=1}^{n} c_i\,\overline{c_j}\,\varphi(g_i \circ g_j^{-1}) = \sum_{i,j=1}^{n} c_i\,\overline{c_j}\,\mathrm{tr}\left[A^* U(g_i \circ g_j^{-1},\lambda)A\right]$$

$$= \left\| A^* \sum_{i=1}^{n} U(g_i,\lambda) \right\|_{HS}^2 \geq 0$$

because

$$U(g_i \circ g_j^{-1},\lambda) = U(g_i,\lambda)\,U^*(g_j,\lambda)\,.$$

And hence $\varphi(g;\lambda)$ is a function of positive type. Moreover, by the same reasoning, if $f(g)$ is any functions for which $\hat{f}(\lambda)$ is a Hermitian positive definite matrix, then $f(g)$ will be a positive definite function. And, according to Hewitt and Ross [9, 11] (p. 683 Lemma D.12), if A and B are both positive definite matrices, then so is their product. This has implications regarding the positivity of the convolution of positive functions on a group.

In particular, if $\rho_t(g) = \rho(g;t)$ is the solution to a driftless diffusion equation with Dirac-delta initial conditions, then the Fourier-space solution is written as

$$\hat{\rho}(\lambda;t) = \exp\left[\frac{1}{2}\sum_{i,j=1}^{dim(G)} D_{ij} u(E_i,\lambda) u(E_j,\lambda)\right]\,,$$

which is Hermitian positive definite, and hence $\rho_t(g)$ is a real-valued positive definite function for each value of $t \in \mathbb{R}_{\geq 0}$. Moreover,[4]

$$\rho_t(g) = \rho_t(g^{-1})\,.$$

It is not difficult to show that given two symmetric functions, $\rho_1(g) \doteq \rho_{t_1}(g;D_1)$ and $\rho_2(g) \doteq \rho_{t_2}(g;D_2)$, that

$$(\rho_1 * \rho_2)(g) = (\rho_2 * \rho_1)(g^{-1})\,.$$

Though this does not imply that $(\rho_1 * \rho_2)(g)$ is symmetric, it is easy to show that $(\rho_1 * \rho_2 * \rho_1)(g)$ is symmetric.

Moreover, if $f : G \to \mathbb{R}_{\geq 0}$ is a pdf which is not symmetric, it is not difficult to show that

$$f'(g) \doteq \frac{f(g) + f(g^{-1})}{2}$$

[4] Here the dependence on $D = [D_{ij}]$ has been suppressed, but really $\rho_t(g) = \rho_t(g;D)$.

and

$$f''(g) \doteq \frac{f(g)f(g^{-1})}{(f*f)(e)}$$

are symmetric pdfs.

For any positive definite symmetric pdf, the Fourier transform is a positive definite Hermitian matrix because

$$\begin{aligned}\hat{\rho}(\lambda) &= \int_G \rho(g)\, U(g^{-1},\lambda)\, dg = \int_G \rho(g^{-1})\, U(g^{-1},\lambda)\, dg \\ &= \int_G \rho(g)\, U(g,\lambda)\, dg = \int_G \rho(g)\, U^*(g^{-1},\lambda)\, dg = \hat{\rho}^*(\lambda)\,.\end{aligned}$$

From positive definiteness, it is possible to write

$$\hat{\rho}(\lambda) = \exp H(\lambda)$$

where H is Hermitian, though not necessarily positive definite.

Moreover, every special unitary matrix can be expressed as the exponential of a skew-Hermitian matrix, and even more than that, if $g = \exp X$, then the IUR matrix $U(g,\lambda)$ can be computed as

$$U(g,\lambda) \;=\; \exp Z(\log(g),\lambda)$$

where

$$Z(X,\lambda) = \sum_{i=1}^{dim(G)} x_i u(E_i,\lambda) = -Z^*(X,\lambda)\,.$$

In analogy with the way that the exponential map for a matrix Lie group is simply the matrix exponential defined by the Taylor series, here and throughout this work the logarithm is the matrix logarithm defined by its Taylor series.

In this light, the Fourier inversion formula has in it the evaluation of

$$\mathrm{tr}\left[\exp H \exp Z\right]\,,$$

and the evaluation of $(\rho_1 * \rho_2)(e)$ has in it

$$\mathrm{tr}\left[\exp H_1 \exp H_2\right]\,.$$

Also, in the evaluation of probability densities for diffusion processes with drift, it is desirable to find approximations of the form

$$\mathrm{tr}\left[\exp(H+Z)\right] \;\approx\; \mathrm{tr}\left[\exp H' \exp Z')\right]\,.$$

For these reasons, there are connections between harmonic analysis on unimodular Lie groups and trace inequalities.

1.4 Trace Inequalities

In the Fourier reconstruction formula for a diffusion process on a Lie group, the trace of the product of exponentials of two matrices is computed. It is therefore relevant to consider: (1) when can the product of two exponentials be simplified; (2) even when the product cannot be simplified, it would be useful to determine when the trace operation has the effect of simplifying the result. For example, if A and B are bandlimited matrices and $\mathrm{tr}(e^A e^B) \approx \mathrm{tr}(\exp(A+B))$ then computing the eigenvalues of $A+B$, exponentiating each eigenvalue, and summing potentially could be much faster than directly exponentiating the matrices, and then taking the trace of the product. The statements that follow therefore may have some relevance to the rapid evaluation of the Fourier inversion formula for diffusion processes on Lie groups.

1.4.1 Generalized Golden-Thompson Inequalities

For $n \times n$ Hermitian matrices A and B, the *Golden-Thompson inequality* [42, 40, 43] is

$$\varphi(e^A e^B) \geq \varphi(e^{A+B}) \tag{1.9}$$

where φ is one of a large number of so-called spectral functions. For the case when $\varphi(\cdot) = \mathrm{tr}(\cdot)$ (which is the case of primary interest in this chapter) this was proven in [42], and generalized in [51].

The Thompson Conjecture [39]: If H and K are Hermitian matrices, there exist unitary matrices U and V dependent on H and K such that

$$e^{iH} e^{iK} = e^{i(UHU^* + VKV^*)}. \tag{1.10}$$

The So-Thompson Conjecture [38]: If H and K are Hermitian matrices, there exist unitary matrices U and V dependent on H and K such that

$$e^{H/2} e^{K} e^{H/2} = e^{UHU^* + VKV^*}. \tag{1.11}$$

Interestingly, such conjectures have been proven [47] using techniques associated with random walks on symmetric space of Lie groups [46], thereby bringing the problem back to the domain of interest in this chapter.

In [41], Cohen et al. prove that *spectral matrix functions*[5], $\varphi : \mathbb{C}^{n\times n} \to \mathbb{C}$ (including the trace) satisfy the inequality

[5] These are functions that depend only the eigenvalues of a matrix, and are therefore invariant under similarity tranformations.

$$\varphi(e^{(A+A^*)/2}e^{(B+B^*)/2}) \geq |\varphi(e^{A+B})| \tag{1.12}$$

when the following condition holds:

$$\varphi([XX^*]^s) \geq |\varphi(X^{2s})| \quad \forall\, X \in \mathbb{C}^{n\times n} \quad \text{for} \quad s = 1, 2, ...$$

An interesting corollary to (1.12) is that if A is skew-Hermitian, and $B \in \mathbb{C}^{n\times n}$, then [41]:

$$\varphi(e^{(B+B^*)/2}) \geq |\varphi(e^{A+B})|. \tag{1.13}$$

Bernstein proved the following general statement for $A \in \mathbb{C}^{n\times n}$ [49]:

$$\operatorname{tr}(e^{A^*}e^{A}) \leq \operatorname{tr}(e^{A^*+A}) \tag{1.14}$$

Inequalities involving functions of products of exponentials have a long history (see e.g., [50, 45]) and remain an area of active investigation. A few recent papers include [52, 53, 54].

1.4.2 Matrix Inequalities from Systems Theory

Another sort of matrix inequality that may be useful would be extensions of results that comes from systems theory. For example, it is known that if $A, B \in \mathbb{R}^{n\times n}$ and $B = B^T > 0$ then [64]

$$\lambda_n(\hat{A})\operatorname{tr}(B) \leq \operatorname{tr}(AB) \leq \lambda_1(\hat{A})\operatorname{tr}(B) \tag{1.15}$$

where $\hat{A} = (A + A^T)/2$ and the eigenvalues are ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$. This has been tightened by Fang et al. [55]:

$$\lambda_n(\hat{A})\operatorname{tr}(B) - \lambda_n(B)[n\lambda_n(\hat{A}) - \operatorname{tr}(A)] \leq \operatorname{tr}(AB) \leq \lambda_1(\hat{A})\operatorname{tr}(B) - \lambda_n(B)[n\lambda_1(\hat{A}) - \operatorname{tr}(A)] \tag{1.16}$$

Additional modifications have been made by Park [63].

Under the same conditions on A and B, Komaroff and Lasserre independently derived the inequality [56, 59]:

$$\sum_{i=1}^{n} \lambda_i(\hat{A})\lambda_{n-i+1}(B) \leq \operatorname{tr}(AB) \leq \sum_{i=1}^{n} \lambda_i(\hat{A})\lambda_i(B) \tag{1.17}$$

and Lasserre tightened this with the result [60]:

$$f(-\epsilon) \leq \operatorname{tr}(AB) \leq f(\epsilon) \quad \forall\, \epsilon > 0 \tag{1.18}$$

where

$$f(\epsilon) \doteq \frac{1}{\epsilon}\sum_{i=1}^{n}[\lambda_i(B + \epsilon\hat{A})\lambda_i(B) - \operatorname{tr}(B^2)].$$

Apparently, when $A \in \mathbb{C}^{n\times n}$ and $B = B^* \in \mathbb{C}^{n\times n}$ (1.17) holds with the substitutions $A^T \to A^*$ and $\mathrm{tr}(AB) \to \mathrm{Re}[\mathrm{tr}(AB)]$ [62]. Generalized formulas for products of arbitrary real matrices have been made recently [61, 62, 65].

1.4.3 Classical Matrix Inequalities

In a sense, inequalities of the form in the previous section can be traced back to work done as early as the 1930s. Mirsky [69] attributes the following result for arbitrary $A, B \in \mathbb{C}^{n\times n}$ to a 1937 paper by John von Neumann:

$$|\mathrm{tr}(AB)| \leq \sum_{i=1}^{n} \mu_i(A)\mu_i(B) \tag{1.19}$$

where $\mu_1(A) \geq \mu_2(A) \geq \cdots \geq \mu_n(A)$ are the singular values of A.

Hoffman and Wieland [57] states that for $n \times n$ normal matrices A and B (i.e., $A^*A = AA^*$ and $B^*B = BB^*$) permutations $\pi, \sigma \in \Pi_n$ can be found such that

$$\sum_{i=1}^{n} |\lambda_i(A) - \lambda_{\pi(i)}|^2 \leq \|A - B\|^2 \leq \sum_{i=1}^{n} |\lambda_i(A) - \lambda_{\sigma(i)}|^2 \tag{1.20}$$

where $\|A\|^2 = \mathrm{tr}(AA^*)$ is the Frobenius norm. For generalizations see [58]. In particular, Richter [72] and Mirsky [67] have shown that if A and B are both $n \times n$ Hermitian matrices,

$$\sum_{i=1}^{n} \lambda_i(A)\lambda_{n+1-i}(B) \leq \mathrm{tr}(AB) \leq \sum_{i=1}^{n} \lambda_i(A)\lambda_i(B) \tag{1.21}$$

and Marcus [66] showed that for normal matrices A and B, there exist permutations π and σ for which

$$\sum_{i=1}^{n} \lambda_i(A)\lambda_{\pi(i)}(B) \leq \mathrm{Re}[\mathrm{tr}(AB)] \leq \sum_{i=1}^{n} \lambda_i(A)\lambda_{\sigma(i)}(B) \tag{1.22}$$

1.4.4 The Arithmetic-Mean-Geometric-Mean (AM-GM) Inequality

The *arithmetic-geometric-mean inequality* states that the arithmetic mean of a set of positive real numbers is always less than the geometric mean of the same set of numbers:

$$\frac{1}{n}\sum_{i=1}^{n}\lambda_i \geq \prod_{i=1}^{n}\lambda_i^{\frac{1}{n}}. \tag{1.23}$$

This fact can be used to derive many useful inequalities. For example, Steele [76] uses the AM-GM inequality to derive a reverse Cauchy-Schwarz inequality of the form

$$\left(\sum_{k=1}^{n} a_k^2\right)^{\frac{1}{2}} \left(\sum_{k=1}^{n} b_k^2\right)^{\frac{1}{2}} \leq \frac{m+M}{2\sqrt{mM}}\sum_{k=1}^{n} a_k b_k \tag{1.24}$$

where $\{a_k\}, \{b_k\} \subset \mathbb{R}_{>0}$ and $0 < m \leq a_k/b_k \leq M < \infty$ for all $k \in \{1, ..., n\}$.

It is no coincidence that the numbers in (1.23) are denoted as λ_i because when they are interpreted as the eigenvalues of a positive definite Hermitian matrix, $A = A^* > 0$, and so

$$\frac{1}{n}\mathrm{tr}(A) \geq |A|^{\frac{1}{n}}. \tag{1.25}$$

This is useful for bounding the trace of the matrix exponential of a not-necessarily-positive-definite Hermitian matrix, $H = H^*$, since $A = \exp H = A^* > 0$ can be substituted into (1.25). The determinant-trace equality for the matrix exponential $\det(\exp(H)) = e^{\mathrm{tr}(H)}$ then gives $|\exp H|^{\frac{1}{n}} = |e^{\mathrm{tr}(H)}|^{\frac{1}{n}}$ and so

$$\frac{1}{n}\mathrm{tr}(\exp H) \geq e^{\mathrm{tr}(H)/n}. \tag{1.26}$$

Though (1.25) is more fundamental than (1.26), the latter is directly useful in studying properties of diffusion processes on Lie groups.

It is also interesting to note that (1.25) generalizes in several ways. For example, if

$$\mu_p(\lambda_1, ..., \lambda_n) \doteq \left(\frac{1}{n}\sum_{i=1}^{n}\lambda_i^p\right)^{\frac{1}{p}}$$

then the AM-GM inequality can be stated as

$$\lim_{p\to 0}\mu_p(\lambda_1, ..., \lambda_n) \leq \mu_1(\lambda_1, ..., \lambda_n)$$

and more generally,

$$\mu_p(\lambda_1, ..., \lambda_n) \leq \mu_q(\lambda_1, ..., \lambda_n) \qquad p < q. \tag{1.27}$$

For each fixed choice of $\lambda_1, ..., \lambda_n$, the function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ defined by $f(p) = \mu_p(\lambda_1, ..., \lambda_n)$ is an increasing function.

When $p = -1$, the *harmonic mean*

$$\mu_{-1}(\lambda_1, ..., \lambda_n) = n \cdot \left[\sum_{i=1}^{n} \frac{1}{\lambda_i}\right]^{-1}$$

results, and (1.27) for $p = -1$ and $q = 0$ implies that

$$n \cdot \left[\sum_{i=1}^{n} \frac{1}{\lambda_i}\right]^{-1} \leq \left[\prod_{i=1}^{n} \lambda_i\right]^{\frac{1}{n}}.$$

The implication of this for positive definite Hermitian matrices is that

$$n \cdot \mathrm{tr}[A^{-1}] \leq [\det A]^{\frac{1}{n}} \iff \mathrm{tr}A \leq \frac{1}{n}[\det A]^{-\frac{1}{n}}. \tag{1.28}$$

One generalization of (1.23) is the *weighted AM-GM inequality*

$$\frac{1}{\alpha}\sum_{i=1}^{n} \alpha_i \lambda_i \geq \left(\prod_{i=1}^{n} \lambda_i^{\alpha_i}\right)^{\frac{1}{\alpha}} \quad \text{where} \quad \alpha = \sum_{i=1}^{n} \alpha_i \ , \ \alpha_i \in \mathbb{R}_{>0}. \tag{1.29}$$

Another generalization is *Ky Fan's inequality* [75]:

$$\frac{\frac{1}{n}\sum_{i=1}^{n} \lambda_i}{\frac{1}{n}\sum_{i=1}^{n}(1-\lambda_i)} \geq \frac{\prod_{i=1}^{n} \lambda_i^{\frac{1}{n}}}{\prod_{i=1}^{n}(1-\lambda_i)^{\frac{1}{n}}} \tag{1.30}$$

which holds for $0 \leq \lambda_i \leq \frac{1}{2}$. If the numbers λ_i are viewed as the eigenvalues of a matrix, A, then Ky Fan's inequality can be written as

$$\frac{\mathrm{tr}A}{\mathrm{tr}(\mathbb{I}-A)} \geq \frac{|A|^{\frac{1}{n}}}{|\mathbb{I}-A|^{\frac{1}{n}}} \quad \text{or} \quad \mathrm{tr}A \geq \frac{n}{1+|A|^{-\frac{1}{n}}|\mathbb{I}-A|^{\frac{1}{n}}}. \tag{1.31}$$

Of course this statement should then be restricted to those matrices that have real eigenvalues that obey $0 \leq \lambda_i(A) \leq \frac{1}{2}$.

1.4.5 Consequences for Harmonic Analysis and Diffusion Processes

If $\rho_{D^{(k)}}(g,t)$ denotes the solution to the driftless diffusion equation on G with diffusion coefficients $D_{ij}^{(k)}$, subject to initial conditions $\rho_{D^{(k)}}(g,0) = \delta(g)$, then from the Golden-Thompson inequality

$$(\rho_{D^{(1)}} * \rho_{D^{(2)}})(e;t) \ \geq \ \rho_{D^{(1)}+D^{(2)}}(e;t)\,. \tag{1.32}$$

Alternatively, using the fact that $B^{\frac{1}{2}}AB^{\frac{1}{2}}$ is Hermitian positive definite whenever A and B are, the trace of the product $\mathrm{tr}[AB] = \mathrm{tr}[B^{\frac{1}{2}}AB]$ can be bounded using (1.25), resulting in

$$\frac{1}{n}\mathrm{tr}(AB) \geq |A|^{\frac{1}{n}}|B|^{\frac{1}{n}}. \tag{1.33}$$

This can then be used to bound $(\rho_{D^{(1)}} * \rho_{D^{(2)}})(e;t)$ from below as well. But it can also be used in a different way. Given a diffusion with drift, the Fourier matrices will be of the form $\exp(H+Z)$ where $H = H^*$ and $Z = -Z^*$. Then the convolution of two diffusions with drifts being the negative of each other will be $\exp(H+Z)\exp(H-Z)$, which is Hermitian, and hence

$$\frac{1}{n}\mathrm{tr}(\exp(H+Z)\exp(H-Z)) \geq |\exp(H+Z)|^{\frac{1}{n}}|\exp(H-Z)|^{\frac{1}{n}}. \tag{1.34}$$

Then from the determinant-trace equality, we can simplify

$$|\exp(H+Z)| = e^{\mathrm{tr}(H+Z)} \quad \text{and} \quad |\exp(H-Z)| = e^{\mathrm{tr}(H-Z)}\,,$$

thereby giving that

$$\frac{1}{n}\mathrm{tr}(\exp(H+Z)\exp(H-Z)) \geq e^{\frac{2}{n}\mathrm{tr}(H)}. \tag{1.35}$$

These will be demonstrated in Section 1.7.

In the case when $G = \mathbb{R}^n$, covariances add under convolution and for a diffusion $\Sigma^{(k)} = D^{(k)}t$, and so

$$(\rho_{D^{(1)}} * \rho_{D^{(2)}})(\mathbf{0};t) = \frac{1}{(2\pi t)^{n/2}\left|D^{(1)} + D^{(2)}\right|^{\frac{1}{2}}} = \rho_{D^{(1)}+D^{(2)}}(\mathbf{0};t)\,.$$

This begs the question of how to define and propagate covariances on a unimodular Lie group, and what relationships may exist with Fisher information. Inequalities relating Fisher information and entropy are reviewed in Section 1.5, followed by definitions of covariance in Section 1.6 and the relationship be Fisher information and covariance.

1.5 Inequalities Involving Fisher Information and Diffusion Processes

This section connects trace inequalities with Fisher information and the rate of entropy increase under a diffusion process. The results presented here are an abridged version of those presented in [36].

1.5.1 Rate of Increase of Entropy under Diffusion

The entropy of a pdf on a Lie group is defined in (1.2) If $f(g,t)$ is a pdf that satisfies a diffusion equation (regardless of the details of the initial conditions) then some interesting properties of $S_f(t)$ can be studied. In particular, if $\dot{S}_f = dS_f/dt$, then differentiating under the integral sign gives

$$\dot{S}_f = -\int_G \left\{ \frac{\partial f}{\partial t} \log f + \frac{\partial f}{\partial t} \right\} dg.$$

But from the properties of a diffusion equation,

$$\int_G \frac{\partial f}{\partial t}\, dg = \frac{d}{dt} \int_G f(g,t)\, dg = 0,$$

and so the second term in the above braces integrates to zero.

Substitution of

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

into the integral for $\dot{S}_f$ gives

$$\begin{aligned}
\dot{S}_f &= -\int_G \left\{ \frac{1}{2} \sum_{i,j=1}^{n} D_{ij} \tilde{E}_i^r \tilde{E}_j^r f - \sum_{k=1}^{n} h_k \tilde{E}_k^r f \right\} \log f \, dg \\
&= -\frac{1}{2} \sum_{i,j=1}^{n} D_{ij} \int_G (\tilde{E}_i^r \tilde{E}_j^r f) \log f \, dg - \sum_{k=1}^{n} h_k \int_G (\tilde{E}_k^r f) \; \log f \, dg \\
&= \frac{1}{2} \sum_{i,j=1}^{n} D_{ij} \int_G (\tilde{E}_j^r f)(\tilde{E}_i^r \log f) \, dg + \sum_{k=1}^{n} h_k \int_G f \, (\tilde{E}_k^r \log f) \, dg \\
&= \frac{1}{2} \sum_{i,j=1}^{n} D_{ij} \int_G \frac{1}{f} (\tilde{E}_j^r f)(\tilde{E}_i^r f) \, dg + \sum_{k=1}^{n} h_k \int_G \tilde{E}_k^r f \, dg \\
&= \frac{1}{2} \sum_{i,j=1}^{n} D_{ij} \int_G \frac{1}{f} (\tilde{E}_j^r f)(\tilde{E}_i^r f) \, dg \\
&\geq 0
\end{aligned}$$

1.5.2 The Generalized de Briujn Identity

This section generalizes the de Bruijn identity, in which entropy rates are related to Fisher information.

Theorem 1. *Let $f_{D,\mathbf{h},t}(g) = f(g,t;D,\mathbf{h})$ denote the solution of the diffusion equation (1.4) with constant $\mathbf{h} = [h_1, ..., h_n]^T$ subject to the initial condition $f(g,0;D,\mathbf{h}) = \delta(g)$. Then for any well-behaved pdf $\alpha(g)$,*

$$\frac{d}{dt}S(\alpha * f_{D,\mathbf{h},t}) = \frac{1}{2}\mathrm{tr}[DF^r(\alpha * f_{D,\mathbf{h},t})]. \tag{1.36}$$

Proof. It is easy to see that the solution of the diffusion equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\sum_{i,j=1}^{n} D_{ij}\tilde{E}_i^r \tilde{E}_j^r \rho - \sum_{k=1}^{n} h_k \tilde{E}_k^r \rho \tag{1.37}$$

subject to the initial conditions $\rho(g,0) = \alpha(g)$ is simply $\rho(g,t) = (\alpha * f_{D,\mathbf{h},t)})(g)$. This follows because all derivatives "pass through" the convolution integral for $\rho(g,t)$ and act on $f_{D,\mathbf{h},t}(g)$.

Taking the time derivative of $S(\rho(g,t))$ gives

$$\frac{d}{dt}S(\rho) = -\frac{d}{dt}\int_G \rho(g,t)\log\rho(g,t)\,dg = -\int_G \left\{\frac{\partial \rho}{\partial t}\log\rho + \frac{\partial \rho}{\partial t}\right\}dg. \tag{1.38}$$

Using (1.37), the partial with respect to time can be replaced with Lie derivatives. But

$$\int_G \tilde{E}_k^r \rho\, dg = \int_G \tilde{E}_i^r \tilde{E}_j^r \rho\, dg = 0,$$

so the second term on the right side of (1.38) completely disappears. Using the integration-by-parts formula[6]

$$\int_G f_1\, \tilde{E}_k^r f_2\, dg = -\int_G f_2\, \tilde{E}_k^r f_1\, dg,$$

with $f_1 = \log\rho$ and $f_2 = \rho$ then gives

$$\begin{aligned}\frac{d}{dt}S(\alpha * f_{D,\mathbf{h},t}) &= \frac{1}{2}\sum_{i,j=1}^{n} D_{ij}\int_G \frac{1}{\alpha * f_{D,\mathbf{h},t}}\tilde{E}_j^r(\alpha * f_{D,\mathbf{h},t})\tilde{E}_i^r(\alpha * f_{D,\mathbf{h},t})\,dg \\ &= \frac{1}{2}\sum_{i,j=1}^{n} D_{ij}F_{ij}^r(\alpha * f_{D,\mathbf{h},t}) \;=\; \frac{1}{2}\mathrm{tr}\left[D\,F^r(\alpha * f_{D,\mathbf{h},t})\right].\end{aligned}$$

The implication of this is that

$$S(\alpha * f_{D,\mathbf{h},t_2}) - S(\alpha * f_{D,\mathbf{h},t_1}) = \frac{1}{2}\int_{t_1}^{t_2} \mathrm{tr}\left[DF^r(\alpha * f_{D,\mathbf{h},t})\right]dt.$$

[6] There are no surface terms because, like the circle and real line, each coordinate in the integral either wraps around or goes to infinity.

1.6 Mean, Covariance, and Their Propagation Under Convolution

This section reviews concepts of mean and covariance for unimodular matrix Lie groups, and how they propagate under convolution. In these definitions, the concepts of Lie-theoretic exponential and logarithm play central roles. For a matrix Lie group, G, with corresponding Lie algebra, $\mathcal{G}$, the exponential map

$$\exp : \mathcal{G} \longrightarrow G$$

simply can be viewed as the matrix exponential defined by the Taylor series. In general, this map is neither surjective nor injective. However, it is possible to characterize the largest path-connected subset $\mathcal{G}^{\circ} \subset \mathcal{G}$ for which the image $G^{\circ} \doteq \exp(\mathcal{G}^{\circ}) \subset G$ has a well-defined inverse map

$$\log : G^{\circ} \longrightarrow \mathcal{G}\,.$$

This is also simply the matrix logarithm defined by its Taylor series.

For $SO(3)$, $SE(2)$, and $SE(3)$ which are three of the most common unimodular matrix Lie groups encountered in applications, the exponential map is surjective and G and G° differ only by a set of measure zero.

In what follows, it is assumed that all probability density functions $f : G \longrightarrow \mathbb{R}_{\geq 0}$ are either supported in G°, or that

$$\int_G f(g)\,dg = \epsilon + \int_{G^{\circ}} f(g)\,dg$$

where ϵ is an inconsequential probability. With this in mind, it becomes possible to blur the difference between G and G°.

1.6.1 Defining Mean

At least three different definitions for the mean of a pdf on a unimodular Lie group exist in the literature. The definitions reviewed here are all in the context of matrix-Lie-theoretic language which grew out of the author's applied work [87, 88]. For similar definitions expressed in differential-geometric terms see [89, 90, 91].

Directly generalizing the definition

$$\mathbf{m} = \int_{\mathbb{R}^n} \mathbf{x}\, f(\mathbf{x})\, d\mathbf{x}$$

to a Lie group is problematic because $\int_G g\, f(g)\, dg$ is not an element of the group. However, it is possible to define $m_0 \in G$ such that

$$\log m_0 = \int_G \log g \, f(g) \, dg \, .$$

Alternatively,

$$\int_{\mathbb{R}^n} (\mathbf{x} - \mathbf{m}) \, f(\mathbf{x}) \, d\mathbf{x} = \mathbf{0}$$

generalizes to searching for $m_1 \in G$ such that

$$\int_G \log \left(m_1^{-1} \circ g \right) \, f(g) \, dg \; = \; \mathbb{O}.$$

Thirdly,

$$\mathbf{m} = \underset{\mathbf{y} \in \mathbb{R}^n}{\operatorname{argmin}} \int_{\mathbb{R}^n} (\mathbf{x} - \mathbf{y})^2 \, f(\mathbf{x}) \, d\mathbf{x}$$

generalizes as

$$m_2 = \underset{h \in G}{\operatorname{argmin}} \int_G \left\| \log \left(h^{-1} \circ g \right) \right\|^2 \, f(g) \, dg \, .$$

In general, no two of m_0, m_1, and m_2 are equal. However, in practice for distributions that are concentrated, they are quite close to each other. That said, if $f(g) = \rho(g)$ is a symmetric function, then all three reduce to the identity element, and hence are equal in this special case.

Though m_0 seems simple and straight forward, it has the undesirable property that shifting a symmetric pdf as $\rho(\mu^{-1} \circ g)$ does not automatically shift the mean from e to μ. m_2 has the problem that the norm $\| \cdot \|$ requires a choice of metric, and for noncompact unimodular Lie groups, a bi-invariant metric generally does not exist. Therefore, conjugating by an arbotrary $a \in G$ a symmetric pdf as $\rho(a^{-1} \circ g \circ a)$, which in the Euclidean setting would leave the mean fixed at e, results in a change to the value of the mean m_2 which depends on a.

In contrast, the mean m_1 shifts naturally with shifts of the pdf because

$$\int_G \log \left(m_1^{-1} \circ g \right) \, \rho(\mu^{-1} \circ g) \, dg \; = \; \int_G \log \left(m_1^{-1} \circ \mu \circ h \right) \, \rho(h) dh.$$

hence $m_1^{-1} \circ \mu = e$, or $m_1 = \mu$. Under conjugation of the pdf, the appearance of log linearly in the definition of m_1 means that

$$\int_G \log \left(m_1^{-1} \circ g \right) \rho(a^{-1} \circ g \circ a) \, dg = \int_G \log \left(m_1^{-1} \circ a \circ h \circ a^{-1} \right) \rho(h) \, dh \; = \; \mathbb{O}$$

can be written as

$$\int_G a^{-1} \log\left(m_1^{-1} \circ a \circ h \circ a^{-1}\right) \, a\, \rho(h)\, dh \;=\; a^{-1}\, \mathbb{O}\, a = \mathbb{O}.$$

But since

$$a^{-1} \log(g)\, a = \log(a^{-1} \circ g \circ a)\,,$$

then the mean m_1 of the conjuagated pdf will be the conjugated mean. The implication of this general result in the special case when ρ is symmetric is

$$a^{-1} \circ m_1^{-1} \circ a = e \implies m_1 = e,$$

giving the desirable property of invariance of the mean of a symmetric function under conjugation.

For these reasons, m_1 is chosen here (and in the author's previous work) as the best definition of the mean, and this is what will be used henceforth, and denoted as μ. The value of μ can be obtained numerically with an iterative procedure using m_0 as the initial starting point.

1.6.2 Defining Covariance

Previously the concepts of $\log : G^{\circ} \to \mathcal{G}$ and $\vee : \mathcal{G} \to \mathbb{R}^n$ where defined. The composition of these maps is defined as

$$\log^{\vee} G^{\circ} \to \mathbb{R}^n\,.$$

That is, for any $g \in G^{\circ}$, $\log^{\vee}(g) \in \mathbb{R}^n$.

One way to define the covariance of pdf on a unimodular Lie group G is [87, 88, 36, 35]

$$\Sigma \doteq \int_G \log^{\vee}(\mu^{-1} \circ g)[\log^{\vee}(\mu^{-1} \circ g)]^T f(g)\, dg\,. \tag{1.39}$$

This definition is natural as a generalization of the concept of covariance in Euclidean space when the pdf of interest is relatively concentrated. Then, for example, a Gaussian distribution can be defined as

$$f(g;\mu,\Sigma) \;\doteq\; \frac{1}{(2\pi)^{d/2}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}[\log^{\vee}(\mu^{-1} \circ g)]^T \Sigma^{-1} \log^{\vee}(\mu^{-1} \circ g)\right)\,.$$

This definition makes sense when the tails decay to negligible values inside a ball around μ for which the exponential and logarithm maps form a bijective pair. Otherwise, the topological properties of G become relevant.

Alternative definitions of scalar variance can be found in [36, 33, 8]. If the covariance as defined in (1.39) has been computed for pdfs f_1 and f_2, a convenient and accurate approximation for the covariance of $f_1 * f_2$ is known

[88, 35]. This is known as a covariance propagation formula. In contrast, the scalar definitions in [33, 8] have exact propagation formulas, but these quantities do not have the form or properties that are usually associated with covariance of pdfs on Euclidean space.

An altogether different way to define covariance that does not involve any approximation is to recognize that for a Gaussian distribution with the mean serving as the statistic, the Cramér-Rao Bound becomes the equality

$$\Sigma_{gaussian} = F_{gaussian}^{-1}, \tag{1.40}$$

and since a Gaussian distribution with $\mu = e$ solves a driftless diffusion equation subject to Dirac delta initial conditions, it is possible to define a kind of covariance for such processes by computing the Fisher information and using (1.40). By generalization, an alternative definition of covariance can be taken as

$$\Sigma' \doteq F^{-1}.$$

The exact properties of this definition under convolution are unknown.

The covariance propagation formula for (1.39) involves the concept of the adjoint matrix, $Ad(g)$. This concept is reviewed in the following section.

1.6.3 The Adjoint Operators ad and Ad

Given $X, Y \in \mathcal{G}$, "little ad" operator is defined as

$$ad_Y(X) \doteq [Y, X] = YX - XY,$$

and "big Ad" is

$$Ad_g(X) \doteq gXg^{-1}.$$

Here "ad" is short for "adjoint". Both of these are linear in X. That is, for arbitrary $c_1, c_2 \in \mathbb{R}$ and $X_1, X_2 \in \mathbb{R}^{n \times n}$

$$ad_Y(c_1X_1 + c_2X_2) = c_1 ad_Y(X_1) + c_2 ad_Y(X_2)$$

and

$$Ad_g(c_1X_1 + c_2X_2) = c_1 Ad_g(X_1) + c_2 Ad_g(X_2).$$

Sometimes Ad_g is written as $Ad(g)$ and ad_Y is written as $ad(Y)$. It turns out that these are related as

$$Ad(\exp(Y)) = \exp(ad(Y)).$$

By introducing a basis for the Lie algebra of a Lie group, it is possible to express the Ad and ad operators as square matrices of the same dimension as the group. The distinction between operators and matrices can sometimes be confusing, which is why, for example, the matrices of $ad(X)$ and $Ad(A)$ are written as $[ad(X)]$ and $[Ad(A)]$ in [35, 36] where

$$[ad(X)]_{ij} = (E_i, ad(X)E_j) \quad \text{and} \quad [Ad(A)]_{ij} = (E_i, Ad(A)E_j)$$

computed using the inner product $(\cdot, \cdot)$.

1.6.4 Covariance Propagation

Given two pdfs with mean and covariance specified, i.e., $f_{(\mu_i,\Sigma_i)}(g)$ for $i = 1, 2$, one would like to be able to write expressions for μ_3, Σ_3 such that

$$f_{(\mu_1,\Sigma_1)} * f_{(\mu_2,\Sigma_2)} = f_{(\mu_3,\Sigma_3)}.$$

In the case when $G = \mathbb{R}^n$, the result is simply $\mathbf{m}_3 = \mathbf{m}_1 + \mathbf{m}_2$, and $\Sigma_3 = \Sigma_1 + \Sigma_2$. This result is nonparametric. That is, it does not require the pdfs to have a specific form, such as a Gaussian.

For general unimodular Lie groups, there is no simple exact formula. However, when the pdfs are concentrated, i.e., both $\|\Sigma_i\|$ are small, then it is possible to write [87]

$$\mu_3 \approx \mu_1 \circ \mu_2 \;\text{ and }\; \Sigma_3 \approx \Sigma_1^{\mu_2} + \Sigma_2.$$

where

$$\Sigma_1^{\mu_2} \doteq [Ad_{\mu_2^{-1}}]\Sigma_1[Ad_{\mu_2^{-1}}]^T.$$

A higher-order approximation of the form [88]

$$\Sigma_3 \approx \Sigma_1^{\mu_2} + \Sigma_2 + \Phi(\Sigma_1^{\mu_2}, \Sigma_2) \tag{1.41}$$

is also possible, but in many applications $\Phi(\Sigma_1^{\mu_2}, \Sigma_2)$ is negligible because it depends quadratically on the elements of $\Sigma_1^{\mu_2}$ and Σ_2.

If $f_{(\mu,\Sigma)}(g) = f(g; \mu, \Sigma)$ denotes a Gaussian distribution in exponential coordinates on a d-dimensional Lie group, it will be of the form

$$f(g;\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}[\log^\vee(\mu^{-1}\circ g)]^T \Sigma^{-1} \log^\vee(\mu^{-1}\circ g)\right).$$

Then

$$f_{(\mu_1,\Sigma_1)} * f_{(\mu_2,\Sigma_2)} \approx f_{(\mu_3,\Sigma_3)}\,. \tag{1.42}$$

The quality of these approximations, as well as those that are even more accurate, have been studied in [35]. This is also a nonparametric result.

In the case when the distributions are more spread out, it is possible to compute the covariance in a different way using the group Fourier transform using the convolution theorem. For example, if $\Sigma_i = D^{(i)} t$ and

$$f_{(\mu_i,\Sigma_i)}(g) \doteq \rho_{\Sigma_i}(\mu_i^{-1} \circ g; t)\,,$$

and since

$$\hat{f}_i(\lambda) = U(\mu_i, \lambda) \exp\left(\sum_{j,k} \sigma_{jk}^{(i)} E_j E_k\right),$$

from the convolution theorem it is possible to write the Fourier version of (1.42) as

$$U(\mu_2, \lambda) \exp\left(\sum_{j,k} \sigma_{jk}^{(2)} E_j E_k\right) U(\mu_1, \lambda) \exp\left(\sum_{j,k} \sigma_{jk}^{(1)} E_j E_k\right) \approx U(\mu_3, \lambda) \exp\left(\sum_{j,k} \sigma_{jk}^{(3)} E_j E_k\right).$$

which can then be substituted in the reconstruction formula to reproduce (1.42), which produces an approximate expression involving traces, which is of a different type than the trace inequalities studied previously in the literature.

1.7 Examples

This section illustrates ideas presented earlier in this paper on the Heisenberg and rotation groups.

1.7.1 The Heisenberg Group, $\mathcal{H}(3)$

The Heisenberg group, $H(3)$, is defined by elements of the form

$$g(\alpha, \beta, \gamma) = \begin{pmatrix} 1 & \alpha & \beta \\ 0 & 1 & \gamma \\ 0 & 0 & 1 \end{pmatrix} \quad \text{where} \quad \alpha, \beta, \gamma \in \mathbb{R} \tag{1.43}$$

and the operation of matrix multiplication. Therefore, the group law can be viewed in terms of parameters as

$$g(\alpha_1, \beta_1, \gamma_1) g(\alpha_2, \beta_2, \gamma_2) = g(\alpha_1 + \alpha_2, \beta_1 + \beta_2 + \alpha_1 \gamma_2, \gamma_1 + \gamma_2).$$

The identity element is the identity matrix $g(0, 0, 0)$, and the inverse of an arbitrary element $g(\alpha, \beta, \gamma)$ is

$$g^{-1}(\alpha, \beta, \gamma) = g(-\alpha, \alpha\gamma - \beta, -\gamma).$$

1.7.1.1 Lie Algebra and Exponential Map

Basis elements for the Lie algebra are

$$E_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad E_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad E_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \tag{1.44}$$

A linear mapping between the Lie algebra spanned by this basis with $\mathbb{R}^3$ is defined by $E_i^\vee = \mathbf{e}_i$.

The Lie bracket is defined as $[E_i, E_j] = E_iE_j - E_jE_i = -[E_j, E_i]$, and so, as is always the case, $[E_i, E_i] = \mathbb{O}$. For these particular basis elements,

$$[E_1, E_2] = [E_2, E_3] = \mathbb{O} \quad \text{and} \quad [E_1, E_3] = E_2.$$

In addition, all double brackets involving the first two listed above are also zero,

$$[E_i, [E_1, E_2]] = [E_i, [E_2, E_3]] = \mathbb{O} \text{ for } i = 1, 2, 3$$

From these, and the bilinearity of the Lie bracket, it follows that for arbitrary

$$X = \sum_i x_i E_i \text{ and } Y = \sum_j x_j E_j$$

that

$$[X, Y] = \sum_{i,j} x_i y_j [E_i, E_j] = (x_1 y_3 - x_3 y_1) E_2 \,. \tag{1.45}$$

If the inner product for the Lie algebra spanned by these basis elements is defined as $(X, Y) = \mathrm{tr}(XY^T)$, then this basis is orthonormal: $(E_i, E_j) = \delta_{ij}$.

The group $H(3)$ is nilpotent because $(x_1E_1 + x_2E_2 + x_3E_3)^n = 0$ for all $n \geq 3$. As a result, the matrix exponential is a polynomial in the coordinates $\{x_i\}$:

$$\exp \begin{pmatrix} 0 & x_1 & x_2 \\ 0 & 0 & x_3 \\ 0 & 0 & 0 \end{pmatrix} = g(x_1, x_2 + \frac{1}{2}x_1x_3, x_3). \tag{1.46}$$

The parametrization in (1.43) can be viewed as the following product of exponentials:

$$g(\alpha, \beta, \gamma) = g(0, \beta, 0)g(0, 0, \gamma)g(\alpha, 0, 0) = \exp(\beta E_2)\exp(\gamma E_3)\exp(\alpha E_1).$$

The logarithm is obtained by solving for each x_i as a function of α, β, γ. By inspection this is $x_1 = \alpha$, $x_3 = \gamma$ and $x_2 = \beta - \alpha\gamma/2$. Therefore,

$$\log g(\alpha, \beta, \gamma) = \begin{pmatrix} 0 & \alpha & \beta - \alpha\gamma/2 \\ 0 & 0 & \gamma \\ 0 & 0 & 0 \end{pmatrix}.$$

1.7.1.2 Adjoint Matrices for $\mathcal{H}(3)$

The adjoint matrix, defined by $[Ad(g)]\mathbf{x} = (gXg^{-1})^{\vee}$, is computed by evaluating

$$\begin{pmatrix} 1 & \alpha & \beta \\ 0 & 1 & \gamma \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & x_1 & x_2 \\ 0 & 0 & x_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\alpha & \alpha\gamma - \beta \\ 0 & 1 & -\gamma \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & x_1 & -\gamma x_1 + x_2 + \alpha x_3 \\ 0 & 0 & x_3 \\ 0 & 0 & 0 \end{pmatrix}.$$

Therefore,

$$(gXg^{-1})^{\vee} = \begin{pmatrix} x_1 \\ -\gamma x_1 + x_2 + \alpha x_3 \\ x_3 \end{pmatrix} \quad \text{and} \quad [Ad(g(\alpha, \beta, \gamma))] = \begin{pmatrix} 1 & 0 & 0 \\ -\gamma & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix}.$$

The fact that $\det[Ad(g)] = 1$ for all $g \in G$ indicates that this group is unimodular. This fact is independent of the parametrization. It can also be shown that for $X = \sum_{i=1}^{3} x_i E_i$ that

$$[Ad(\exp X)] = \begin{pmatrix} 1 & 0 & 0 \\ -x_3 & 1 & x_1 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.47}$$

1.7.2 Bi-Invariant Integration Measure

The Jacobian matrices for this group can be computed in either parametrization. In terms of α, β, γ,

$$\frac{\partial g}{\partial \alpha} = E_1; \quad \frac{\partial g}{\partial \beta} = E_2; \quad \frac{\partial g}{\partial \gamma} = E_3.$$

A straightforward calculation then gives

$$g^{-1}\frac{\partial g}{\partial \alpha} = E_1; \quad g^{-1}\frac{\partial g}{\partial \beta} = E_2; \quad g^{-1}\frac{\partial g}{\partial \gamma} = E_3 - \alpha E_2.$$

Therefore

$$J_r(\alpha, \beta, \gamma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\alpha \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad J_l(\alpha, \beta, \gamma) = \begin{pmatrix} 1 & 0 & 0 \\ -\gamma & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.48}$$

Then $|J_l(\alpha, \beta, \gamma)| = |J_r(\alpha, \beta, \gamma)| = 1$ and bi-invariant integration measure expressed in these coordinates is simply

$$dg = d\alpha d\beta d\gamma\,.$$

In exponential coordinates

$$J_r(\mathbf{x}) = \begin{pmatrix} 1 & 0 & 0 \\ x_3/2 & 1 & -x_1/2 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad J_l(\mathbf{x}) = \begin{pmatrix} 1 & 0 & 0 \\ -x_3/2 & 1 & x_1/2 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.49}$$

and

$$dg = dx_1 dx_2 dx_3 \,.$$

1.7.3 Covariance Propagation and the EPI for $\mathcal{H}(3)$

From (1.45)the only nonzero term in the second-order covariance propagation formula from [88, 35] is

$$\frac{1}{4}[X,Y]^{\vee}\left([X,Y]^{\vee}\right)^T = \frac{1}{4}(x_1 y_3 - x_3 y_1)^2 \mathbf{e}_2 \mathbf{e}_2^T .$$

Hence, for $\mathcal{H}(3)$, the second-order term in (1.41), which results from integrating the above as was done in [88], becomes

$$\Phi(\Sigma_1^{\mu_2}, \Sigma_2) \;=\; a\mathbf{e}_2\mathbf{e}_2^T$$

where

$$a = \frac{1}{4}\left(\sigma_{11}^{(1)}\sigma_{33}^{(2)} - \sigma_{13}^{(1)}\sigma_{31}^{(2)} - \sigma_{31}^{(1)}\sigma_{13}^{(2)} + +\sigma_{33}^{(1)}\sigma_{11}^{(2)}\right) \;\geq\; 0$$

with $\sigma_{ij}^{(1)} = \mathbf{e}_i^T \Sigma_1^{\mu_2}\mathbf{e}_j$ and $\sigma_{ij}^{(2)} = \mathbf{e}_i^T \Sigma_2 \mathbf{e}_j$.

Moreover, from the matrix identity for $A = A^T > 0$,

$$\det(A + \mathbf{u}\mathbf{v}^T) = (1 + \mathbf{v}^T A^{-1}\mathbf{u}) \det A \,,$$

it follows that

$$\det(\Sigma_{(1*2)}) = \det(\Sigma_1^{\mu_2} + \Sigma_2 + a\mathbf{e}_2\mathbf{e}_2^T) = (1 + a\mathbf{e}_2^T(\Sigma_1^{\mu_2} + \Sigma_2)^{-1}\mathbf{e}_2) \det(\Sigma_1^{\mu_2} + \Sigma_2) \,,$$

where $\Sigma_1^{\mu_2} \;=\; Ad_{\mu_2^{-1}} \Sigma_1 Ad_{\mu_2^{-1}}^T$. Consequently, from the Euclidean EPI and the unimodularity of G, which implies that

$$\det(\Sigma_1^{\mu_2}) = \det(\Sigma_1) \,,$$

it is not difficult to see that

$$\det(\Sigma_{(1*2)})^{\frac{1}{\dim(G)}} \;\geq\; \det(\Sigma_1^{\mu_2} + \Sigma_2)^{\frac{1}{\dim(G)}} \;\geq\; \det(\Sigma_1)^{\frac{1}{\dim(G)}} + \det(\Sigma_2)^{\frac{1}{\dim(G)}} .$$

That is, the entropy power inequality for pdfs from diffusion processes on $\mathcal{H}(3)$ follows in the small time limit from the classical EPI, and it is less restrictive than in the Euclidean case.

1.7.4 The Case of $SO(3)$

The group of rotations of three-dimensional space has elements that are 3×3 special orthogonal matrices, i.e., those satisfying

$$RR^T = \mathbb{I} \quad \text{and} \quad \det(R) = +1\,.$$

That is, they satisfy $RR^T = \mathbb{I}$ and $\det R = +1$. It is easy to see that closure of these propeties under multiplication is satisfied because

$$(R_1R_2)^T(R_1R_2) = R_2^T R_1^T R_1 R_2 = R_2^T R_2 = \mathbb{I}$$

and

$$\det(R_1R_2) = \det(R_1)\det(R_2) = 1 \cdot 1 = 1.$$

1.7.4.1 The Lie Algebra

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

$$X = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix} = \sum_{i=1}^{3} x_i E_i\,. \tag{1.50}$$

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

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

For $SO(3)$ the adjoint matrices are

$$[Ad(R)] = R \quad \text{and} \quad [ad(X)] = X.$$

Furthermore,

$$[X, Y]^\vee = \mathbf{x} \times \mathbf{y}.$$

1.7.4.2 Exponential and Logarithm

It is well known that the exponential map $\exp : so(3) \to SO(3)$ is related to Euler's Theorem as

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

where $\theta \in [0, \pi]$ is the angle of rotation around the axis $\mathbf{n} \in S^2$, with N being the associated skew-symmetric matrix. Then $X = \theta N$ and $\mathbf{x} = \theta \mathbf{n}$. It is convenient to limit $\theta \in [0, \pi]$ and to allow $\mathbf{n}$ to take any value in the unit sphere, S^2. Moreover,

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

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 $X = \theta N = \log R$, it follows that

$$\log(R) \;=\; \frac{\cos^{-1}\left[\frac{\mathrm{tr}(R)-1}{2}\right](R - R^T)}{\sqrt{3 - (\mathrm{tr}(R))^2 + 2\mathrm{tr}(R)}}\,. \tag{1.51}$$

This expression breaks down when $\theta = \pi$, which defines a set of measure zero, and hence is inconsequential when evaluating the logarithm under an integral.

1.7.4.3 Invariant Integration Measure

Two common ways to parameterize rotations are using the matrix exponential $R = \exp X$ and using Euler angles such as $R = R_3(\alpha)R_1(\beta)R_3(\gamma)$ where $0 \leq \alpha, \gamma \leq 2\pi$ and $0 \leq \beta \leq \pi$.

Relatively simple analytical expressions were derived by Park [97] for the Jacobian J_l when $R = \exp X$ as

$$J_l(\mathbf{x}) = \mathbb{I} + \frac{1 - \cos\|\mathbf{x}\|}{\|\mathbf{x}\|^2} X + \frac{\|\mathbf{x}\| - \sin\|\mathbf{x}\|}{\|\mathbf{x}\|^3} X^2 \tag{1.52}$$

The corresponding Jacobian J_r and its inverse are [35, 36]

$$J_r(\mathbf{x}) = \mathbb{I} - \frac{1 - \cos\|\mathbf{x}\|}{\|\mathbf{x}\|^2} X + \frac{\|\mathbf{x}\| - \sin\|\mathbf{x}\|}{\|\mathbf{x}\|^3} X^2$$

In terms of ZXZ Euler angles,

$$J_l(\alpha,\beta,\gamma) = [\mathbf{e}_3, R_3(\alpha)\mathbf{e}_1, R_3(\alpha)R_1(\beta)\mathbf{e}_3] = \begin{pmatrix} 0 & \cos\alpha & \sin\alpha\sin\beta \\ 0 & \sin\alpha & -\cos\alpha\sin\beta \\ 1 & 0 & \cos\beta \end{pmatrix}. \tag{1.53}$$

and

$$J_r = R^T J_l = [R_3(-\gamma)R_1(-\beta)\mathbf{e}_3, R_3(-\gamma)\mathbf{e}_1, \mathbf{e}_3] = \begin{pmatrix} \sin\beta\sin\gamma & \cos\gamma & 0 \\ \sin\beta\cos\gamma & -\sin\gamma & 0 \\ \cos\beta & 0 & 1 \end{pmatrix}. \tag{1.54}$$

From this we see that

$$dR = \frac{2(1-\cos\|\mathbf{x}\|)}{\|\mathbf{x}\|^2}\, dx_1 dx_2 dx_3 = \sin\beta\, d\alpha d\beta d\gamma\,.$$

From these it can be shown that

$$\int_{SO(3)} dR = 8\pi^2\,.$$

1.7.4.4 Fourier Series

For $SO(3)$, irreducible unitary representations (IURs) [35] 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. The explicit forms of these matrices when R is expressed in Euler angles are well known in Physics as the Wigner-D functions [92, 93, 94, 95, 96]

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

$$\hat{f}^l_{mn} = \int_{SO(3)} f(A) U^l_{mn}(A^{-1})\, dA\,. \tag{1.55}$$

The following orthogonality relation holds

$$\int_{SO(3)} U^l_{mn}(A)\overline{U^s_{pq}(A)}\, dA = \frac{1}{2l+1}\delta_{ls}\delta_{mp}\delta_{nq} \tag{1.56}$$

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

$$f(A) = \sum_{l=0}^{\infty}(2l+1)\sum_{m=-l}^{l}\sum_{n=-l}^{l} \hat{f}^l_{mn} U^l_{nm}(A)\,, \tag{1.57}$$

which results from the completeness relation

$$\sum_{l=0}^{\infty}(2l+1)\sum_{m=-l}^{l}\sum_{n=-l}^{l}U_{mn}^{l}(R^{-1})U_{nm}^{l}(A)=\delta(R^{-1}A)\,. \tag{1.58}$$

Another way to write (1.57) is

$$f(A)=\sum_{l=0}^{\infty}(2l+1)\mathrm{trace}\left[\hat{f}^{l}\,U^{l}(A)\right]\,. \tag{1.59}$$

1.7.5 Diffusions on $SO(3)$

A diffusion process on $SO(3)$ commonly encountered in applications is of the form

$$\frac{\partial f}{\partial t}=\frac{1}{2}\sum_{i,j=1}^{3}D_{ij}\tilde{E}_i\tilde{E}_j f+\sum_{k=1}^{3}d_k\tilde{E}_k f\,. \tag{1.60}$$

By expanding the PDF in the PDE in (1.60) into a Fourier series on $SO(3)$, the solution can be obtained once we know how the differential operators X_i^R transform the matrix elements $U_{m,n}^{l}(A)$. Explicitly,

$$\tilde{E}_1U_{mn}^{l}=\frac{1}{2}c_{-n}^{l}U_{m,n-1}^{l}-\frac{1}{2}c_{n}^{l}U_{m,n+1}^{l}; \tag{1.61}$$

$$\tilde{E}_2U_{mn}^{l}=\frac{1}{2}ic_{-n}^{l}U_{m,n-1}^{l}+\frac{1}{2}ic_{n}^{l}U_{m,n+1}^{l}; \tag{1.62}$$

$$\tilde{E}_3U_{mn}^{l}=-inU_{mn}^{l}; \tag{1.63}$$

where $c_n^l=\sqrt{(l-n)(l+n+1)}$ for $l\geq|n|$ and $c_n^l=0$ otherwise. From this definition it is clear that $c_k^k=0$, $c_{-(n+1)}^l=c_n^l$, $c_{n-1}^l=c_{-n}^l$, and $c_{n-2}^l=c_{-n+1}^l)$.

By repeated application of these rules, it can be shown that [35]

$$\mathcal{F}\left(\frac{1}{2}\sum_{i,j=1}^{3}D_{ij}\tilde{E}_i\tilde{E}_j f+\sum_{i=1}^{3}d_i\tilde{E}_i f\right)_{mn}^{l}=\sum_{k=\max(-l,m-2)}^{\min(l,m+2)}\mathcal{A}_{m,k}^{l}\hat{f}_{k,n}^{l},$$

where

$$\mathcal{A}_{m,m+2}^{l}=\left[\frac{(D_{11}-D_{22})}{8}+\frac{i}{4}D_{12}\right]c_{m+1}^{l}c_{-m-1}^{l};$$

$$\mathcal{A}_{m,m+1}^{l}=\left[\frac{(2m+1)}{4}(D_{23}-iD_{13})+\frac{1}{2}(d_1+id_2)\right]c_{-m-1}^{l};$$

$$\mathcal{A}^l_{m,m} = \left[-\frac{(D_{11}+D_{22})}{8}(c^l_{-m}c^l_{m-1} + c^l_m c^l_{-m-1}) - \frac{D_{33}m^2}{2} - id_3 m\right];$$

$$\mathcal{A}^l_{m,m-1} = \left[\frac{(2m-1)}{4}(D_{23}+iD_{13}) + \frac{1}{2}(-d_1 + id_2)\right] c^l_{m-1};$$

$$\mathcal{A}^l_{m,m-2} = \left[\frac{(D_{11}-D_{22})}{8} - \frac{i}{4}D_{12}\right] c^l_{-m+1}c^l_{m-1};$$

Hence, application of the $SO(3)$-Fourier transform to (1.60) and corresponding initial conditions reduces (1.60) to a set of linear time-invariant ODEs of the form

$$\frac{d\hat{f}^l}{dL} = \mathcal{A}^l \hat{f}^l \quad \text{with} \quad \hat{f}^l(0) = \mathbb{I}_{2l+1}. \tag{1.64}$$

Here $\mathbb{I}_{2l+1}$ is the $(2l+1)\times(2l+1)$ identity matrix and the banded matrix $\mathcal{A}^l$ are of the following form for $l = 0, 1, 2, 3$:

$$\mathcal{A}^0 = \mathcal{A}^0_{0,0} = 0; \quad \mathcal{A}^1 = \begin{pmatrix} \mathcal{A}^1_{-1,-1} & \mathcal{A}^1_{-1,0} & \mathcal{A}^1_{-1,1} \\ \mathcal{A}^1_{0,-1} & \mathcal{A}^1_{0,0} & \mathcal{A}^1_{0,1} \\ \mathcal{A}^1_{1,-1} & \mathcal{A}^1_{1,0} & \mathcal{A}^1_{1,1} \end{pmatrix};$$

$$\mathcal{A}^2 = \begin{pmatrix} \mathcal{A}^2_{-2,-2} & \mathcal{A}^2_{-2,-1} & \mathcal{A}^2_{-2,0} & 0 & 0 \\ \mathcal{A}^2_{-1,-2} & \mathcal{A}^2_{-1,-1} & \mathcal{A}^2_{-1,0} & \mathcal{A}^2_{-1,1} & 0 \\ \mathcal{A}^2_{0,-2} & \mathcal{A}^2_{0,-1} & \mathcal{A}^2_{0,0} & \mathcal{A}^2_{0,1} & \mathcal{A}^2_{0,2} \\ 0 & \mathcal{A}^2_{1,-1} & \mathcal{A}^2_{1,0} & \mathcal{A}^2_{1,1} & \mathcal{A}^2_{1,2} \\ 0 & 0 & \mathcal{A}^2_{2,0} & \mathcal{A}^2_{2,1} & \mathcal{A}^2_{2,2} \end{pmatrix};$$

$$\mathcal{A}^3 = \begin{pmatrix} \mathcal{A}^3_{-3,-3} & \mathcal{A}^3_{-3,-2} & \mathcal{A}^3_{-3,-1} & 0 & 0 & 0 & 0 \\ \mathcal{A}^3_{-2,-3} & \mathcal{A}^3_{-2,-2} & \mathcal{A}^3_{-2,-1} & \mathcal{A}^3_{-2,0} & 0 & 0 & 0 \\ \mathcal{A}^3_{-1,-3} & \mathcal{A}^3_{-1,-2} & \mathcal{A}^3_{-1,-1} & \mathcal{A}^3_{-1,0} & \mathcal{A}^3_{-1,1} & 0 & 0 \\ 0 & \mathcal{A}^3_{0,-2} & \mathcal{A}^3_{0,-1} & \mathcal{A}^3_{0,0} & \mathcal{A}^3_{0,1} & \mathcal{A}^3_{0,2} & 0 \\ 0 & 0 & \mathcal{A}^3_{1,-1} & \mathcal{A}^3_{1,0} & \mathcal{A}^3_{1,1} & \mathcal{A}^3_{1,2} & \mathcal{A}^3_{1,3} \\ 0 & 0 & 0 & \mathcal{A}^3_{2,0} & \mathcal{A}^3_{2,1} & \mathcal{A}^3_{2,2} & \mathcal{A}^3_{2,3} \\ 0 & 0 & 0 & 0 & \mathcal{A}^3_{3,1} & \mathcal{A}^3_{3,2} & \mathcal{A}^3_{3,3} \end{pmatrix}.$$

The solution to (1.64) is then of the form of a matrix exponential:

$$\hat{f}^l(L) = e^{L\mathcal{A}^l}. \tag{1.65}$$

Since $\mathcal{A}^l$ is a band-diagonal matrix for $l > 1$, the matrix exponential can be calculated much more efficiently (either numerically or symbolically) for large values of l than for general matrices of dimension $(2l+1)\times(2l+1)$.

Given the explicit forms provided above, (1.32)-(1.35) can be verified.

1.7.5.1 Lack of an Entropy-Power Inequality

For all unimodular Lie groups, the EPI holds for concentrated Gaussian pdfs for which the first-order covariance propagation formula from ([87]) holds by application of the Euclidean EPI to Gaussians. However, for compact Lie groups (including the circle and n-torus) the EPI always breaks down. For example, the uniform distribution on the circle, $\rho(\theta) = 1/2\pi$, has entropy $S(\rho) = \log(1/2\pi)$. But since this distribution is stable under convolution, we have that $S(\rho * \rho) = S(\rho)$ and so the EPI cannot hold since $N(\rho * \rho) = N(\rho) < 2 \cdot N(\rho)$. Similarly, unlike for $\mathcal{H}(3)$, the EPI does not hold for $SO(3)$.

1.8 Conclusions

Many inequalities of information theory that are based on probability densities on Euclidean space extend to the case of probabilities on Lie groups. In addition to reviewing appropriate concepts of integration, convolution, partial derivative, Fourier transform, covariance, and diffusion processes on unimodular Lie groups, this paper also presents some new inequalities that extend to this setting those known in the classical Abelian case.

References

1. Chirikjian, G.S., “Degenerate Diffusions and Harmonic Analysis on SE(3): A Tutorial,” in *Stochastic Geometric Mechanics*, (S. Albeverio, A. Cruzeiro, D.Holm, eds.), pp. 77-99, Springer, 2017.
2. Simon, B., *Trace Ideals and Their Applications,* 2^{nd} *ed.*, Mathematical Surveys and Monographs, American Mathematical Society, 2010
3. Chirikjian, G.S., “Information-Theoretic Inequalities on Unimodular Lie Groups,” *Journal of Geometric Mechanics*, 2(2):119 – 158, June 2010.
4. Wang, Y., Zhou, Y., Maslen, D.K., Chirikjian, G.S., “Solving the Phase-Noise Fokker-Planck Equation Using the Motion-Group Fourier Transform,” *IEEE Transactions on Communications*, 54 (5): 868 – 877 May 2006.
5. Zhou, Y., Chirikjian, G.S., “Conformational Statistics of Semi-Flexible Macromolecular Chains with Internal Joints,” *Macromolecules*, 39(5), pp. 1950 – 1960, 2006.
6. Chirikjian, G.S., Kyatkin, A.B.,“An Operational Calculus for the Euclidean Motion Group with Applications in Robotics and Polymer Science,” *J. Fourier Analysis and Applications*, 6: (6) 583 – 606, December 2000.
7. Folland, G.B., *A Course in Abstract Harmonic Analysis*, CRC Press, Boca Raton, FL, 1995.
8. Grenander, U., *Probabilities on Algebraic Structures*, Dover, 2008.
9. Gross, K.I., “Evolution of Noncommutative Harmonic Analysis,” *American Mathematical Monthly*, 85(7): 525 – 548, 1978.
10. Gurarie, D., *Symmetry and Laplacians. Introduction to Harmonic Analysis, Group Representations and Applications*, Elsevier Science Publisher, The Netherlands, 1992. (Dover Edition, 2008).

11. Hewitt, E., Ross, K.A., *Abstract Harmonic Analysis I, and II*, Springer-Verlag, Berlin, 1963 and 1970. (reprinted 1994).
12. Miller, W., Jr., *Lie Theory and Special Functions*, Academic Press, New York, 1968;
13. Miller, W. Jr., "Some Applications of the Representation Theory of the Euclidean Group in Three-Space," *Commun. Pure App. Math.*, Vol. 17, pp. 527-540, 1964.
14. Sugiura, M., *Unitary Representations and Harmonic Analysis*, 2^{nd} edition, North-Holland, Amsterdam, 1990.
15. Taylor, M.E., *Noncommutative Harmonic Analysis*, Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1986.
16. Thangavelu, S., *Harmonic Analysis on the Heisenberg Group*, Birkhäuser, Boston, 1998.
17. Vilenkin, N.Ja. Klimyk, A.U., *Representation of Lie Groups and Special Functions*, Vols. 1-3, Kluwer Academic Publ., Dordrecht, Holland 1991.
18. Vilenkin, N.J., *Special Functions and the Theory of Group Representations*, American Mathematical Society, 1968.
19. Vilenkin, N.J., Akim, E.L., Levin, A.A., "The Matrix Elements of Irreducible Unitary Representations of the Group of Euclidean Three-Dimensional Space Motions and Their Properties," *Dokl. Akad. Nauk SSSR*, Vol. 112, pp. 987-989, 1957 (in Russian).
20. Howe, R., and Tan, E.C., *Non-Abelian Harmonic Analysis*, Springer, 1992.
21. Lang, S., $SL_2(R)$, Addison-Wesley, 1975
22. Harish Chandra, "Spherical Functions on a Semisimple Lie Group II," American Journal of Mathematics 27:569–579, 1960.
23. Jorgenson, J., Lang, S., *Spherical Inversion on* $SL_n(R)$, Springer, 2001.
24. Neuenschwander, D., *Probabilities on the Heisenberg Group: Limit Theorems and Brownian Motion*, Lecture Notes in Mathematics, 1630, Springer-Verlag, Berlin, 1996.
25. Kunze, R., "L_p Fourier Transforms on Locally Compact Unimodular Groups," *Transactions of the American Mathematical Society*, 89: 519 – 540, 1958.
26. Applebaum, D., *Probability on Compact Lie Groups*, Springer, New York, 2014.
27. Beckner, W., "Sharp inequalities and geometric manifolds," *J. Fourier Anal. Appl.* 3 (1997), 825-836.
28. Beckner, W., "Geometric inequalities in Fourier analysis," Essays on Fourier Analysis in Honor of Elias M. Stein, Princeton University Press, 1995, pp. 36-68
29. Blachman, N.M., "The convolution inequality for entropy powers," *IEEE Trans. Inform. Theory*, 11(2): 267271, 1965.
30. Carlen, E.A., "Superadditivity of Fishers Information and Logarithmic Sobolev Inequalities," *Journal Of Functional Analysis*, 101, 194-211 (1991)
31. Cover, T.M., Thomas, J.A., *Elements of Information Theory*, Wiley-Interscience, 2^{nd} ed., Hoboken, NJ, 2006.
32. Dembo, A., Cover, T.M., Thomas, J.A., "Information Theoretic Inequalities," *IEEE Transactions On Information Theory* 37(6):1501-1518 NOV 1991
33. Heyer, H., *Probability Measures on Locally Compact Groups*, Springer-Verlag, New York, 1977.
34. Bhatia, R., *Positive Definite Matrices*, Princeton University Press, 2007.
35. Chirikjian, G.S., Kyatkin, A.B., *Harmonic Analysis for Engineers and Applied Scientists*, Dover, Mineola, NY, 2016.
36. Chirikjian, G.S., *Stochastic Models, Information Theory, and Lie Groups: Volume 2 - Analytic Methods and Modern Applications*, Birkhäuser, Boston, 2011.
37. Bernstein, D.S., *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*, Princeton University Press (February 22, 2005)
38. So, W., Thompson, R.C., "Products of Exponentials of Hermitian and Complex Symmetric Matrices," *Linear and Multilinear Algebra*, 1991, Vol. 29, pp. 225-233
39. Thompson, R.C., "Special cases of a matrix exponential formula," *Linear Algebra And Its Applications*, 107 (1988), 283-292.

40. Thompson, C.J., “Inequalities and Partial Orders on Matrix Spaces,” *Indiana University Mathematics Journal*, 21(5) 1971, pp. 469-480.
41. Cohen, J.E., Friedland, S., Kato, T., Kelly, F.P., “Eigenvalue Inequalities for Products of Matrix Exponentials,” *Linear Algebra And Its Applications* 45:55-95 (1982)
42. Golden, S., “Lower bounds for the Helmholtz function,” *Phys. Rev.* 137:B1127-B1128 (1965).
43. Thompson, C.J., *Mathematical Statistical Mechanics*, Macmillan, New York, 1972; reprint, Princeton University Press, Princeton, 1979, 1992.
44. Trotter, H.F., “On the product of semi-groups of operators,” *Proc. Amer. Math. Soc.*, 10545-551 (1959).
45. Fan, K., “Maximum properties and inequalities for the eigenvalues of completely continuous operators,” *PNAS* 37 (1951) pp. 760-766.
46. Klyachko, A.A., “Random walks on symmetric spaces and inequalities for matrix spectra,” *Linear Algebra Appl.* 319 (2000) 3759.
47. So, W., “The high road to an exponential formula,” *Linear Algebra and its Applications*, 379 (2004) 6975
48. So, W., “Equality Cases In Matrix Exponential Inequalities,” *SIAM J. Matrix Anal. Appl.*, Vol. 13, No. 4, pp. 1154-1158, October 1992
49. Bernstein, D.S., “Inequalities for the trace of matrix exponentials,” *SIAM J. Matrix Anal. Appl.*, 9 (1988), pp. 156-158.
50. Fan, K., “On a theorem of Weyl concerning eigenvalues of linear transformations I,” *Proc. Nat. Acad. Sci. U.S.A.*, 35 (1949), pp. 652-655.
51. Lenard, A., “Generalization of the Golden-Thompson inequality $Tr(e^A e^B) \geq Tr(e^{A+B})$,” *Indiana Univ. Math. J.*, 21 (1971), pp. 457-467.
52. Bebiano, N., da Providência, J. Jr., Lemos, R., “Matrix inequalities in statistical mechanics,” *Linear Algebra and its Applications*, Volume 376, 1 January 2004, Pages 265-273
53. Friedland, S., So, W., “Product of matrix exponentials,” *Linear Algebra Appl.* 196 (1994) 193205.
54. Friedland, S., Porta, B., “The limit of the product of the parameterized exponentials of two operators,” *Journal of Functional Analysis* 210 (2004) 436464
55. Fang, Y., Loparo, K.A., Feng, X., “Inequalities for the Trace of Matrix Product,” *IEEE Transactions On Automatic Control*, 39(12), December 1994, pp. 2489-2490.
56. Komaroff, N., “Bounds on eigenvalues of matrix products with an application to the algebraic Riccati equation,” *IEEE Trans. Autom. Control*, vol. 35, no. 3, pp. 348350, Mar. 1990.
57. Hoffman, A.J., Wielandt, H.W., “The Variation of the Spectrum of a Normal Matrix,” *Duke Math. J.*, 20 (1953), pp. 37-40.
58. Cochran, J.A., Hinds, E.W., “Improved Error Bounds For The Eigenvalues Of Certain Normal Operators,” *SIAM J. Numer. Anal.*, 9(3), September 1972, pp. 446-453.
59. Lasserre, J.B., “A trace inequality for the matrix product,” *IEEE Trans. Automat. Contr.*, vol. 40, pp. 15001501, 1995
60. Lasserre, J.B., “Tight Bounds for the Trace of a Matrix Product,” *IEEE Transactions on Automatic Control*, 42(4) April 1997, pp. 578-581.
61. Wei Xing, Qingling Zhang, and Qiyi Wang “A Trace Bound for a General Square Matrix Product,” *IEEE Transactions on Automatic Control*, 45(8) August 2000, pp. 1563-1565.
62. Fuzhen Zhang and Qingling Zhang Eigenvalue Inequalities for Matrix Product *IEEE Transactions on Automatic Control*, 51(9), September 2006, pp. 1506-1509.
63. Park, P.-G., “On the Trace Bound of a Matrix Product,” *IEEE Transactions on Automatic Control*, 41(12), December 1996, pp. 1799-1802.
64. T. Mori, Comments on A matrix inequality associated with bounds on solutions of algebraic Riccati and Lyapunov equation, IEEE Trans. Automat. Contr., vol. AC-29, p. 1088, Nov. 1988

65. Jianzhou Liu and Lingli He A New Trace Bound for a General Square Matrix Product *IEEE Transactions on Automatic Control*, 52(2), February 2007, pp. 349-352.
66. Marcus, M., "An Eigenvalue Inequality for the Product of Normal Matrices," The American Mathematical Monthly, Vol. 63, No. 3. (Mar., 1956), pp. 173-174.
67. L. Mirsky, "On the trace of matrix products," *Mathematische Nachrichten*, 20 (1959), pp. 171-174.
68. L. Mirsky, "A Note On Normal Matrices," The American Mathematical Monthly, Vol. 63, No. 7. (Aug. - Sep., 1956), p. 479.
69. Mirsky, L., "A Trace Inequality of John von Neumann," *Monatshefte für Mathematik* 79, pp. 303-306 (1975)
70. Reid, R.M., "Some Eigenvalue Properties of Persymmetric Matrices," SIAM Review, Vol. 39, No. 2. (Jun., 1997), pp. 313-316.
71. Schur, I., "Über die charakteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen," *Math. Annalen*, vol. 66, 1909 ,pp. 488-510.
72. Richter, H., "Zur Abschätzung von Matrizennormen," *Mathematische Nachrichten*, 18 (1958), pp. 178-187.
73. Thurston, H.S., "On the characteristic equations of products of square matrices," *The American Mathematical Monthly*, vol. 38, 1931, pp. 322-324.
74. Scott, W.M., "On Characteristic Roots of Matrix Products," *The American Mathematical Monthly*, Vol. 48, No. 3. (Mar., 1941), pp. 201-203.
75. Neuman, E., Sándor, J., "On the Ky Fan inequality and related inequalities I," *Mathematical Inequalities & Applications* 5 (1): 4956, 2002.
76. Steele, J.M., *The Cauchy-Schwarz master class : an introduction to the art of mathematical inequalities*, Cambridge ; New York : Cambridge University Press, 2004.
77. Varopoulos, N.Th., Saloff-Coste, L., Coulhon, T., *Analysis and Geometry on Groups*, Cambridge University Press, 1992.
78. Varopoulos, N.Th., Saloff-Coste, L., Coulhon, T., *Analysis and Geometry on Groups*, Cambridge University Press, 1992.
79. Maslen, D.K., *Fast Transforms and Sampling for Compact Groups* (Ph.D. Dissertation, Department of Mathematics, Harvard University, May 1993).
80. Maslen, D.K., Rockmore, D.N., "Generalized FFTsa survey of some recent results," *DIMACS Series Discrete Math. Theor. Comput. Sci.* 28, 183237 (1997).
81. Bhatia, R., Parthasarathy, K.R., "Positive Definite Functions and Operator Inequalities," *Bull. London Math. Soc.*, Vol. 32, pp. 214-228, 2000.
82. Andruchow, E., Corach, G., Stojanoff, D., "Geometric Operator Inequalities," *Lin. Alg. Appl.*, Vol. 258, pp. 295-310, 1997.
83. Bhatia, R., Kittaneh, F., "On Singular Values of a Product of Operators," *SIAM J. Matrix Anal. Appl.*, Vol. 11, pp. 272-277, 1990.
84. Hardy, G.H., Littlewood, J.E., Pólya, G., *Inequalities*, Cambridge University Press, 1932.
85. Pólya, G., *Isoperimetric inequalities in mathematical physics*, Princeton, Princeton University Press, 1951.
86. Bhatia, R., *Matrix Analysis*, Springer, 1996.
87. Wang, Y., Chirikjian, G.S., "Error Propagation on the Euclidean Group with Applications to Manipulator Kinematics," *IEEE Transactions on Robotics*, 22(4):591 – 602 August 2006.
88. Wang, Y., Chirikjian, G.S., "Nonparametric Second-Order Theory of Error Propagation on the Euclidean Group," *International Journal of Robotics Research*, Vol. 27, No. 1112, November/December 2008, pp. 1258-1273
89. Pennec, X., *L'incertitude dans les problèmes de reconnaissance et de recalage– Applications en imagerie médicale et biologie moléculaire*, (Doctoral dissertation, Ecole Polytechnique X), 1996.

90. Pennec, X., "Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements," *J. Math Imaging and Vision*, 25:127, July 2006.
91. Pennec, X., and Vincent Arsigny, V., "Exponential Barycenters of the Canonical Cartan Connection and Invariant Means on Lie Groups," In Frederic Barbaresco, Amit Mishra, and Frank Nielsen, editors, *Matrix Information Geometry*, pages 123-166. Springer, May 2012.
92. Biedenharn, L.C., Louck, J.D., *Angular Momentum in Quantum Physics*, Encyclopedia of Mathematics and Its Applications, Vol. 8, Cambridge University Press, 1985. (paperback version 2009).
93. Gelfand, I. M., Minlos, R.A., Shapiro, Z.Ya., *Representations of the Rotation and Lorentz Groups and Their Applications*, Macmillan, New York, 1963.
94. Talman, J., *Special Functions*, W. A. Benjamin, Inc., Amsterdam, 1968.
95. Varshalovich, D.A., Moskalev, A.N., Khersonskii, V.K., *Quantum Theory of Angular Momentum*, World Scientific, Singapore, 1988.
96. Wigner, E.P., *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra*, Academic Press, New York, 1959.
97. Park, F.C., *The Optimal Kinematic Design of Mechanisms*, Ph.D. Thesis, Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 1991.