Information-Theoretic Inequalities on Unimodular Lie Groups

Gregory S. Chirikjian
Department of Mechanical Engineering
Johns Hopkins University
gregc@jhu.edu

June 2, 2009

Abstract

Classical inequalities used in information theory such as those of de Bruijn, Fisher, and Kullback carry over from the setting of probability theory on Euclidean space to that of unimodular Lie groups. These are groups that possess integration measures that are invariant under left and right shifts, which means that even in noncommutative cases they share many of the useful features of Euclidean space. In practical engineering terms the rotation group and Euclidean motion group are the unimodular Lie groups of most interest, and the development of information theory applicable to these Lie groups opens up the potential to study problems relating to image reconstruction from irregular or random projection directions, information gathering in mobile robotics, satellite attitude control, and bacterial chemotaxis and information processing. Several definitions are extended from the Euclidean case to that of Lie groups including the Fisher information matrix, and inequalities analogous to those in classical information theory are derived and stated in the form of fifteen small theorems. In all such inequalities, addition of random variables is replaced with the group product, and the appropriate generalization of convolution of probability densities is employed.

1 Introduction

Shannon’s brand of information theory is now more than six decades old, and some of the statistical methods developed by Fisher, Kullback, etc., are even older. Similarly, the study of Lie groups is now more than a century old. Despite their relatively long and roughly parallel history, surprisingly few connections appear to have been made between these two vast fields. One such connection is in the area of ergodic theory [1, 3, 11], where the Boltzmann-Shannon entropy is replaced with topological entropy [39, 43, 51, 77]. Ergodic theory developed in parallel with information theory and remains an active area of research among mathematicians to the current day (see e.g., [83]). Both use concepts of entropy (though these concepts are quite different from each other), and some common treatments have been given over the years (see e.g., [10]). However, it should be noted that some of the cornerstones of information theory such as the de Bruijn inequality, Fisher information, Kullback-Leibler divergence, etc., do not carry over to ergodic theory. And while connections between ergodic theory and Lie groups are quite strong, connections between information theory and Lie groups are
virtually nonexistent. The goal of this paper is therefore to present a unified framework of “information theory on Lie groups.” As such, fifteen small theorems are presented that involve the structure and/or group operation of Lie groups. Unlike extensions of information theory to manifolds, the added structure inherent in Lie groups allow us to draw much stronger parallels with inequalities of classical information theory, such as those presented in [20, 27].

In recent years a number of connections have begun to emerge linking information theory, group theory, and geometry. A cross section of that work is reviewed here, and it is explained how the results of this paper are distinctly different from prior works.

In the probability and statistics literature, the statistical properties of random walks and limiting distributions on Lie groups has been studied extensively by examining the properties of iterated convolutions [28, 35, 53, 61, 62, 68]. The goal in many of these works is to determine the form of the limiting distribution, and the speed of convergence to it. This is a problem closely related to those in information theory. However, to the author’s knowledge concepts such as entropy, Fisher information, Kullback-Leibler divergence, etc., are not used significantly in those analysis. Rather, techniques of harmonic analysis (Fourier analysis) on Lie groups are used, such as the methods described in [24, 32, 33, 37, 70, 71, 73, 75, 80]. Indeed, to the best of the author’s knowledge the only work that uses the concept and properties of information-theoretic (as opposed to topological) entropy on Lie groups is that of Johnson and Suhov [10, 11]. Their goal was to use the Kullback-Leibler divergence between probability density functions on compact Lie groups to study the convergence to uniformity under iterated convolutions, in analogy with what was done by Linnik [51] and Barron [6] in the commutative case. The goal of the present paper is complementary: using some of the same tools, many of the major defined quantities and inequalities of (differential) information theory are extended from $\mathbb{R}^n$ to the context of unimodular Lie groups, which form a broader class of Lie groups than compact ones.

The goal here is to define and formalize probabilistic and information-theoretic quantities that are currently arising in scenarios such as robotics [48, 54, 56, 65, 72, 53, 47, 76], bacterial motion [9, 74], and parts assembly in automated manufacturing systems [13, 25, 44, 60, 69]. The topics of detection, tracking, estimation and control on Lie groups has been studied extensively over the past four decades. For example, see [14, 15, 18, 29, 32, 67, 44, 55, 58, 53, 50, 75] (and references therein). Many of these problems involve probability densities on the group of rigid-body motions. However, rather than focusing only on rigid-body motions, a general information theory on the much broader class of unimodular Lie groups is presented here with little additional effort.

Several other research areas that would initially appear to be related to the present work have received intensive interest. For decades, Amari has developed the concept of information geometry [2] in which the Fisher information matrix is used to define a Riemannian metric tensor on spaces of probability distributions, thereby allowing those spaces to be viewed as Riemannian manifolds. This provides a connection between information theory and differential geometry. However, in information geometry, the probability distributions themselves (such as Gaussian distributions) are defined on a Euclidean space, rather than on a Lie group.

A different kind of connection between information theory and geometry has been established in the context of medical imaging and computer vision in which probability densities on manifolds are analyzed using information-theoretic techniques [59]. However, a manifold generally does not have an associated group operation, and so there is
no natural way to “add” random variables. Relatively recently, Yeung and coworkers have used the structure of finite groups to derive new inequalities for discrete information. While this heavily involves the use of the theory of finite groups, the goal is to derive new inequalities for classical information theory, i.e., that which is concerned with discrete information related to finite sets. For example, see the work of Chan and Yeung [19, 20] and Zhang and Yueng [81]. Li and Chong [49] and Chan [20] have addressed the relationship between group homomorphisms and information inequalities using the Ingleton inequality. In these works, the groups are discrete, and the new inequalities that are derived pertain to classical informational quantities. In contrast, the goal of the current presentation is to extend concepts from information theory to the case where variables “live in” a Lie group.

While on the one hand work that connects geometry and information theory exists, and on the other hand work that connects finite-group theory and information theory exists, very little has been done along the lines of developing information theory on Lie groups, which in addition to possessing the structure of differential manifolds, also are endowed with group operations. Indeed, it would appear that applications such as deconvolution on Lie groups [21] (which can be formulated in an information-theoretic context [79, 46]), and the field of Simultaneous Localization and Mapping (or SLAM) [72] have preceded the development of formal information inequalities that take advantage of the Lie-group structure of rigid-body motions.

This paper attempts to address this deficit with a two-pronged approach: (1) by collecting some known results from the functional analysis literature and reinterpreting them in information-theoretic terms (e.g. Gross’ log-Sobolev inequality on Lie groups); (2) by defining information-theoretic quantities such as entropy, covariance and Fisher information matrix, and deriving inequalities involving these quantities that parallels those in classical information theory.

The remainder of this paper is structured as follows: Section 2 provides a brief review of the theory of unimodular Lie groups and gives several concrete examples (the rotation group, Euclidean motion group, Heisenberg group, and special linear group). An important distinction between information theory on manifolds and that on Lie groups is that the existence of the group operation in the latter case plays an important role. Section 3 defines entropy and relative entropy for unimodular Lie groups and proves some of their properties under convolution and marginalization over subgroups and coset spaces. The concept of the Fisher information matrix for probability densities on unimodular Lie groups is defined in Section 4 and several elementary properties are proven. This generalized concept of Fisher information is used in Section 5 to establish the de Brujin inequality for unimodular Lie groups. Finally, these definitions and properties are combined with recent results by others on log-Sobolev inequalities in Section 6.

2 A Brief Review of Unimodular Lie Groups

Rather than starting with formal definitions, examples of unimodular Lie groups are first introduced, their common features are enumerated, and then their formal properties are enumerated.
2.1 An Introduction to Lie Groups via Examples

Perhaps one reason why there has been little cross-fertilization between the theory of Lie groups and information theory is that the presentation styles in these two fields are very different. Whereas Lie groups belong to pure mathematics, information theory emerged from engineering. Therefore, this section reviews some of the basic properties of Lie groups from a concrete engineering perspective. All of the groups considered are therefore matrix Lie groups.

2.1.1 Example 1: The Rotation Group

Consider the set of $3 \times 3$ rotation matrices

$$SO(3) = \{ R \in \mathbb{R}^{3\times3} | RR^T = I, \det R = +1 \}.$$  

Here $SO(3)$ denotes the set of special orthogonal $3 \times 3$ matrices with real entries. It is easy to verify that this set is closed under matrix multiplication and inversion. That is, $R, R_1, R_2 \in SO(3) \implies R_1R_2, R^{-1} \in SO(3)$. Furthermore, the $3 \times 3$ identity matrix is in this set, and the associative law $R_1(R_2R_3) = (R_1R_2)R_3$ holds, as is true for matrix multiplication in general. This means that $SO(3)$ is a group, and is called the special orthogonal (or rotation) group. Furthermore, it can be reasoned that the nine independent entries in a $3 \times 3$ real matrix are constrained by the orthogonality condition $RR^T = I$ to the point where a three-degree-of-freedom subspace remains. (The condition $\det R = +1$ does not further constrain the dimension of this subspace, though it does limit the discussion to one component of the space defined by the orthogonality condition).

It is common to describe the three free degrees of freedom of the rotation group using parametrizations such as the ZXZ Euler angles:

$$R(\alpha, \beta, \gamma) = R_3(\alpha)R_1(\beta)R_3(\gamma) \quad (1)$$

where $R_i(\theta)$ is a counterclockwise rotation about the $i^{th}$ coordinate axis. Another popular description of 3D rotations or the axis-angle parametrization

$$R(\vartheta, n) = I + \sin \vartheta N + (1 - \cos \vartheta)N^2 \quad (2)$$

where $N$ is the unique skew-symmetric matrix such that $Nx = n \times x$ for any $x \in \mathbb{R}^3$, and $n$ is the unit vector pointing along the axis of rotation and $\times$ is the vector cross product. The “vee and hat” notation

$$N^\vee = n \quad \iff \quad N = \hat{n} \quad (3)$$

is used to describe this relationship. Here $\|n\| = (n \cdot n)^{\frac{1}{2}} = 1$. It can be parameterized in spherical coordinates as $n = n(\phi, \theta)$, and so a parametrization of the form $R = R(\vartheta, \phi, \theta)$ results. The angles $\vartheta, \phi, \theta$ are not the same as the Euler angles $\alpha, \beta, \gamma$.

The group $SO(3)$ is a compact Lie group, and therefore has finite volume. When using Euler angles, volume is computed with respect to the integration measure

$$dR = \frac{1}{8\pi^2} \sin \alpha \, d\alpha \, d\beta \, d\gamma, \quad (4)$$

which when integrated over $0 \leq \alpha, \gamma \leq 2\pi$ and $0 \leq \beta \leq \pi$ gives a value of 1. Indeed, this result was obtained by construction by using the normalization of $8\pi^2$. The same
volume element will take on a different form when using the axis-angle parametrization, in analogy with the way that the volume element in $\mathbb{R}^3$ can be expressed in the equivalent forms $dxdydz$ and $r^2 \sin \theta d\theta dr d\phi$ in Cartesian and spherical coordinates, respectively.

Given any 3-parameter description of rotation, the angular velocity of a rigid body can be obtained from a rotation matrix. Angular velocity in the body-fixed and space-fixed reference frames can be written respectively as

$$\mathbf{\omega}_r = J_r(q)\dot{q} \quad \text{and} \quad \mathbf{\omega}_l = J_l(q)\dot{q}$$

where $q$ is any parametrization (e.g., $q = [\alpha, \beta, \gamma]^T$ or $q = [\theta, \phi, \theta]^T$, where $T$ denotes the transpose of a vector or matrix).

The Jacobian matrices $J_r(q)$ and $J_l(q)$ are computed from the parametrization $R(q)$ and the definition of the $\vee$ operation in (3) as

$$J_l(q) = \left[ \left( \frac{\partial R}{\partial q_1} R^T \right)^\vee, \left( \frac{\partial R}{\partial q_2} R^T \right)^\vee, \left( \frac{\partial R}{\partial q_3} R^T \right)^\vee \right].$$

and

$$J_r(q) = \left[ \left( R^T \frac{\partial R}{\partial q_1} \right)^\vee, \left( R^T \frac{\partial R}{\partial q_2} \right)^\vee, \left( R^T \frac{\partial R}{\partial q_3} \right)^\vee \right].$$

This gives a hint as to why the subscripts $l$ and $r$ are used: if derivatives with respect to parameters appear on the ‘right’ of $R^T$, this is denoted with an $r$, and if they appear on the ‘left’ then a subscript $l$ is used.

Explicitly for the Euler angles,

$$J_l(\alpha, \beta, \gamma) = [e_3, R_3(\alpha)e_1, R_3(\alpha)R_1(\beta)e_3] = \begin{pmatrix} 0 & \cos \alpha & \sin \alpha \sin \beta \\ 0 & \sin \alpha & -\cos \alpha \sin \beta \\ 1 & 0 & \cos \beta \end{pmatrix}$$

and

$$J_r = R^T J_l = [R_3(-\gamma)R_1(-\beta)e_3, R_3(-\gamma)e_1, e_3] = \begin{pmatrix} \sin \beta \sin \gamma & \cos \gamma & 0 \\ \sin \beta \cos \gamma & -\sin \gamma & 0 \\ \cos \beta & 0 & 1 \end{pmatrix}.$$  \hspace{1cm} (5)

Note that

$$|J_l| = |J_r| = \sin \beta$$

gives the factor that appears in the volume element $dR$ in (4). This is not a coincidence.

For any parametrization of $SO(3)$ of the form $R(q)$, the volume element can be expressed as

$$dR = \frac{1}{8\pi^2}|J(q)|dq_1dq_2dq_3$$

where $J(q)$ can be taken to be either $J_r(q)$ or $J_l(q)$. Though these matrices are not equal, their determinants are.

Whereas the set of all rotations together with matrix multiplication forms a noncommutative ($R_1R_2 \neq R_2R_1$ in general) Lie group, the set of all angular velocity vectors $\mathbf{\omega}_r$ and $\mathbf{\omega}_l$ (or more precisely, their corresponding matrices, $\hat{\mathbf{\omega}}_r$ and $\hat{\mathbf{\omega}}_l$) together with the operations of addition and scalar multiplication form a vector space. Furthermore, this vector space is endowed with an additional operation, the cross product $\mathbf{\omega}_1 \times \mathbf{\omega}_2$ (or equivalently the matrix commutator $[\hat{\mathbf{\omega}}_1, \hat{\mathbf{\omega}}_2] = \hat{\mathbf{\omega}}_1\hat{\mathbf{\omega}}_2 - \hat{\mathbf{\omega}}_2\hat{\mathbf{\omega}}_1$). This makes the set...
of all angular velocities a Lie algebra, which is denoted as \( so(3) \) (as opposed to the Lie group, \( \text{SO}(3) \)).

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

\[
X = \begin{pmatrix}
0 & -x_3 & x_2 \\
x_3 & 0 & -x_1 \\
-x_2 & x_1 & 0
\end{pmatrix} = \sum_{i=1}^{3} x_i X_i.
\] (7)

The skew-symmetric matrices \( \{X_i\} \) form a basis for the set of all such \( 3 \times 3 \) skew-symmetric matrices, and the coefficients \( \{x_i\} \) are all real.

Lie algebras and Lie groups are related in general by the exponential map. For matrix Lie groups (which are the only kind of Lie groups that will be discussed here), the exponential map is the matrix exponential function. In this specific case,

\[
\exp : so(3) \rightarrow \text{SO}(3).
\]

It is well known (see [24] for derivation and references) that

\[
R(\mathbf{x}) = e^X = I + \frac{\sin \| \mathbf{x} \|}{\| \mathbf{x} \|} X + \frac{(1 - \cos \| \mathbf{x} \|)}{\| \mathbf{x} \|^2} X^2
\] (8)

where \( \| \mathbf{x} \| = (x_1^2 + x_2^2 + x_3^2)^{\frac{1}{2}} \). Indeed, (8) is simply a variation on (2) with \( \mathbf{x} = \theta \mathbf{n} \).

An interesting and useful fact is that except for a set of measure zero, all elements of \( \text{SO}(3) \) can be captured with the parameters within the open ball defined by \( \| \mathbf{x} \| < \pi \), and the matrix logarithm of any group element parameterized in this range is also well defined. It is convenient to know that the angle of the rotation, \( \vartheta(R) \), is related to the exponential parameters as \( |\vartheta(R)| = \| \mathbf{x} \| \). Furthermore,

\[
\log(R) = \frac{1}{2 \sin \vartheta(R)} \left( R - R^T \right)
\]

where

\[
\vartheta(R) = \cos^{-1} \left( \frac{\text{trace}(R) - 1}{2} \right).
\]

Relatively simple analytical expressions have been derived for the Jacobian \( J_l \) and its inverse when rotations are parameterized as in (8):

\[
J_l(\mathbf{x}) = I + \frac{1 - \cos \| \mathbf{x} \|}{\| \mathbf{x} \|^2} X + \frac{\| \mathbf{x} \| - \sin \| \mathbf{x} \|}{\| \mathbf{x} \|^3} X^2.
\] (9)

The corresponding Jacobian \( J_r \) is calculated as [24]

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

Note that

\[J_l = J_r^T \text{ and } J_l = R J_r.\]

The determinants are

\[| \det(J_l) | = | \det(J_r) | = \frac{2(1 - \cos \| \mathbf{x} \|)}{\| \mathbf{x} \|^2}.\]
2.1.2 Example 2: The Euclidean Motion Group of the Plane

The Euclidean motion group of the plane can be thought of as the set of all matrices of the form

\[
g(x_1, x_2, \theta) = \begin{pmatrix} \cos \theta & -\sin \theta & x \\ \sin \theta & \cos \theta & y \\ 0 & 0 & 1 \end{pmatrix}
\]

(10)

together with the operation of matrix multiplication.

It is straightforward to verify that the form of these matrices is closed under multiplication and inversion, and that \(g(0, 0, 0) = I\), and that it is therefore a group. This is often referred to as the special Euclidean group, and is denoted as \(SE(2)\). Like \(SO(3)\), \(SE(2)\) is three dimensional. However, unlike \(SO(3)\), \(SE(2)\) is not compact. Nevertheless, it is possible to define a natural integration measure for \(SE(2)\) as

\[
dg = dx dy d\theta.
\]

And while \(SE(2)\) does not have finite volume (and so there is no single natural normalization constant such as \(8\pi^2\) in the case of \(SO(3)\)), this integration measure nevertheless can be used to compute probabilities from probability densities.

Note that

\[
g(x, y, \theta) = \exp(xX_1 + yX_2) \exp(\theta X_3)
\]

where

\[
X_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad X_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}; \quad X_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.
\]

These matrices form a basis for the Lie algebra, \(se(2)\). It is convenient to identify these with the natural basis for \(\mathbb{R}^3\) by defining \((X_i)^\vee = e_i\). In so doing, any element of \(se(2)\) can be identified with a vector in \(\mathbb{R}^3\).

The Jacobians for this parametrization are then of the form

\[
J_l = \left[ \left( \frac{\partial g}{\partial x} g^{-1} \right)^\vee, \left( \frac{\partial g}{\partial y} g^{-1} \right)^\vee, \left( \frac{\partial g}{\partial \theta} g^{-1} \right)^\vee \right] = \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}
\]

and

\[
J_r = \left[ \left( g^{-1} \frac{\partial g}{\partial x} \right)^\vee, \left( g^{-1} \frac{\partial g}{\partial y} \right)^\vee, \left( g^{-1} \frac{\partial g}{\partial \theta} \right)^\vee \right] = \begin{pmatrix} 1 & 0 & y \\ 0 & 1 & -x \\ 0 & 0 & 1 \end{pmatrix}.
\]

Note that

\[
|\det(J_l)| = |\det(J_r)| = 1.
\]

This parametrization is not unique, though it is probably the most well-known one.

As an alternative, consider the exponential parametrization \(\exp : se(2) \to SE(2)\):

\[
g(x_1, x_2, x_3) = \exp(x_1X_1 + x_2X_2 + x_3X_3)
\]

\[
= \exp \begin{pmatrix} 0 & -x_3 & x_1 \\ x_3 & 0 & x_2 \\ 0 & 0 & 0 \end{pmatrix}
\]

\[
= \begin{pmatrix} \cos x_3 & -\sin x_3 & \frac{x_2(-1 + \cos x_3) + x_1 \sin x_3}{x_3} \\ \sin x_3 & \cos x_3 & \frac{x_1(1 - \cos x_3) + x_2 \sin x_3}{x_3} \\ 0 & 0 & 1 \end{pmatrix}.
\]
Comparing this with \((\text{10})\) it is clear that \(x_3 = \theta\), but \(x \neq x_1\) and \(y \neq x_2\).

The Jacobians in this exponential parametrization are

\[
J_r = \begin{pmatrix}
\sin x_3 & \cos x_3 - \frac{1}{x_3} & 0 \\
\frac{1}{x_3} \cos x_3 - \frac{x_3}{x_3} \sin x_3 & \frac{1}{x_3} \sin x_3 & 0 \\
x_3 x_1 - x_2 x_3 - x_1 \cos x_3 - x_3 x_1 \sin x_3 & x_1 + x_2 x_3 - x_1 \cos x_3 - x_3 x_1 \sin x_3 & 1 \\
\end{pmatrix}
\]

\[
J_l = \begin{pmatrix}
\sin x_3 & \frac{1}{x_3} \cos x_3 - \frac{x_3}{x_3} \sin x_3 & 0 \\
\frac{x_3}{x_3} \cos x_3 - \frac{x_3}{x_3} \sin x_3 & \frac{x_3}{x_3} \sin x_3 & 0 \\
x_3 x_1 + x_2 x_3 - x_1 \cos x_3 - x_3 x_1 \sin x_3 & -x_1 + x_2 x_3 + x_1 \cos x_3 - x_3 x_1 \sin x_3 & 1 \\
\end{pmatrix}
\]

It follows that

\[|\det(J_l)| = |\det(J_r)| = \frac{2(1 - \cos x_3)}{x_3^2}.\]

### 2.1.3 Example 3: The Heisenberg Group

The Heisenberg group, \(H(1)\), is defined by elements of the form

\[
g(\alpha, \beta, \gamma) = \begin{pmatrix} 1 & \alpha & \beta \\ 0 & 1 & \gamma \\ 0 & 0 & 1 \end{pmatrix}
\]

where \(\alpha, \beta, \gamma \in \mathbb{R}\) (12)

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

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

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

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

Basis elements for the Lie algebra are

\[
X_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad X_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad X_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \quad (13)
\]

The Lie bracket, \([X_i, X_j] = X_i X_j - X_j X_i\), for these basis elements gives

\[\{X_1, X_2\} = [X_2, X_3] = 0 \quad \text{and} \quad [X_1, X_3] = X_2.\]

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

The group \(H(1)\) is nilpotent because \((x_1 X_1 + x_2 X_2 + x_3 X_3)^n = 0\) for all \(n \geq 3\). As a result, the matrix exponential is a polynomial in the coordinates \(\{x_i\}\):

\[
\exp \begin{pmatrix} 0 & x_1 & x_2 \\ 0 & 0 & x_3 \\ 0 & 0 & 0 \end{pmatrix} = g(x_1, x_2 + \frac{1}{2} x_1 x_3, x_3). \quad (14)
\]

The parametrization in \((\text{14})\) can be viewed as the following product of exponentials:

\[
g(\alpha, \beta, \gamma) = g(0, \beta, 0)g(0, 0, \gamma)g(\alpha, 0, 0) = \exp(\beta X_2) \exp(\gamma X_3) \exp(\alpha E_1).
\]
The logarithm is obtained by solving for each \( x_i \) as a function of \( \alpha, \beta, \gamma \). By inspection this is \( x_1 = \alpha, x_3 = \gamma \) and \( x_2 = \beta - \alpha \gamma / 2 \). Therefore,

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

The Jacobian matrices for this group can be computed in either parametrization. In terms of \( \alpha, \beta, \gamma \),

\[
J_r(\alpha, \beta, \gamma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\alpha \\ 0 & 0 & 1 \end{pmatrix}
\quad \text{and} \quad
J_l(\alpha, \beta, \gamma) = \begin{pmatrix} 1 & 0 & 0 \\ -\gamma & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\]

In terms of exponential coordinates,

\[
J_r(x) = \begin{pmatrix} 1 & 0 & 0 \\ x_3/2 & 1 & -x_1/2 \\ 0 & 0 & 1 \end{pmatrix}
\quad \text{and} \quad
J_l(x) = \begin{pmatrix} 1 & 0 & 0 \\ -x_3/2 & 1 & x_1/2 \\ 0 & 0 & 1 \end{pmatrix}.
\]

In both parametrizations \(|\det J_r| = |\det J_l| = 1|.

2.1.4 Example 4: The Special Linear Group

The group \( SL(2, \mathbb{R}) \) consists of all \( 2 \times 2 \) matrices with real entries with determinant equal to unity. In other words, for \( a, b, c, d \in \mathbb{R} \) elements of \( SL(2, \mathbb{R}) \) are of the form

\[
g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}
\quad \text{where} \quad ad - bc = 1.
\]

Subgroups of \( SL(2, \mathbb{R}) \) include matrices of the form

\[
g_1(x) = \exp \left( \begin{pmatrix} x & 0 \\ 0 & -x \end{pmatrix} \right) = \begin{pmatrix} e^x & 0 \\ 0 & e^{-x} \end{pmatrix};
\]

\[
g_2(y) = \exp \left( \begin{pmatrix} 0 & y \\ 0 & 0 \end{pmatrix} \right) = \begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix};
\]

\[
g_3(\theta) = \exp \left( \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix} \right) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.
\]

A basis for the Lie algebra \( sl(2, \mathbb{R}) \) is

\[
X_1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}; \quad X_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}; \quad X_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.
\]

An inner product can be defined in which this basis is orthonormal.

It can be shown that any \( g \in SL(2, \mathbb{R}) \) can be expressed as a product of \( g_1(x), g_2(y), \) and \( g_3(\theta) \). This is called an Iwasawa decomposition of \( SL(2, \mathbb{R}) \).

The above \( g_i \) are not the only subgroups of \( SL(2, \mathbb{R}) \). For example, exponentiating matrices of the form \( \xi \cdot (X_3 + 2X_2) \) results in a subgroup of matrices of the form

\[
g(\xi) = \begin{pmatrix} \cosh \xi & \sinh \xi \\ \sinh \xi & \cosh \xi \end{pmatrix}.
\]
The Iwasawa decomposition allows one to write an arbitrary \( g \in SL(2, \mathbb{R}) \) in the form \[70\]

\[
g = g_1(\theta)g_2(t)g_3(\xi)
\]

where

\[
u_1(\theta) = \exp(\theta X_1) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix};
\]

\[
u_2(t) = \exp(t X_2) = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix};
\]

\[
u_3(\xi) = \exp(\xi (X_3 - X_1)) = \begin{pmatrix} 1 & \xi \\ 0 & 1 \end{pmatrix}.
\]

In this parametrization the right Jacobian is

\[
J_r(\theta, t, \xi) = \frac{1}{2} \begin{pmatrix} e^{-2t} + e^{2t}(1 + \xi^2) & -2e^{2t}\xi & e^{2t} - e^{-2t}(1 + e^{4t}\xi^2) \\ -2\xi & 2 & 2\xi \\ 1 & 0 & 1 \end{pmatrix}.
\]

The left Jacobian is

\[
J_l(\theta, t, \xi) = \frac{1}{2} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2\cos 2\theta & 2\sin 2\theta \\ -e^{2t} & -e^{2t}\sin 2\theta & e^{2t}\cos 2\theta \end{pmatrix}.
\]

It is easy to verify that

\[|\det(J_r(\theta, t, \xi))| = |\det(J_l(\theta, t, \xi))| = \frac{1}{2} e^{2t}.\]

Hence, \( SL(2, \mathbb{R}) \) is unimodular (which means the determinants of the left and right Jacobians are the same).

2.2 Generalizations

Whereas several low-dimensional examples of Lie groups were presented to make the discussion concrete, a vast variety of different kinds of Lie groups exist. For example, the same constraints that were used to define \( SO(3) \) relative to \( \mathbb{R}^{3\times3} \) can be used to define \( SO(n) \) from \( \mathbb{R}^{n\times n} \). The result is a Lie group of dimension \( n(n-1)/2 \) and has a natural volume element \( dR \). Similarly, the Euclidean motion group generalizes as all \( (n+1) \times (n+1) \) matrices of the form

\[
g = \begin{pmatrix} R & t \\ 0^T & 1 \end{pmatrix} = \left( \begin{array}{c|c} \mathbb{I} & t \\ \hline 0^T & 1 \end{array} \right) \left( \begin{array}{c|c} R & 0 \\ \hline 0^T & 1 \end{array} \right)
\]

resulting in \( SE(n) \) having dimension \( n(n+1)/2 \) and natural volume element \( dg = dRdt \) where \( t \in \mathbb{R}^n \) and \( dt = dt_1dt_2\cdots dt_n \) is the natural integration measure for \( \mathbb{R}^n \). The following subsections briefly review the general theory of Lie groups that will be relevant when defining information-theoretic inequalities.
2.2.1 Exponential, Logarithm, and Vee Operation

In general an n-dimensional real matrix Lie algebra is defined by a basis consisting of real matrices \( \{ X_i \} \) for \( i = 1, \ldots, n \) that is closed under the matrix commutator. That is, \( [X_i, X_j] = \sum_{k=1}^{n} C_{ij}^{k} X_k \) for some real numbers \( \{ C_{ij}^{k} \} \), which are called the structure constants of the Lie algebra.

In a neighborhood around the identity of the corresponding Lie group, the parametrization

\[
g(x_1, \ldots, x_n) = \exp X \quad \text{where} \quad X = \sum_{i=1}^{n} x_i X_i
g(18)
\]

is always valid in a region around the identity in the corresponding Lie group. And in fact, for the examples discussed, this parametrization is good over almost the whole group, with the exception of a set of measure zero.

The logarithm map

\[
\log g(x) = X
\]

(which is the inverse of the exponential) is valid except on this set of measure zero. It will be convenient in the analysis to follow to identify a vector \( x \in \mathbb{R}^n \) as

\[
x = (\log g)^\vee \quad \text{where} \quad (X_i)^\vee = e_i.
g(19)
\]

Here \( \{ e_i \} \) is the natural basis for \( \mathbb{R}^n \).

In terms of quantities that have been defined in the examples, the adjoint matrices \( Ad \) and \( ad \) are the following matrix-valued functions:

\[
Ad(g) = J_l J_r^{-1} \quad \text{and} \quad ad(X) = \log Ad(e^X).
g(20)
\]

The dimensions of these square matrices is the same as the dimension of the Lie group, which can be very different than the dimensions of the matrices that are used to represent the elements of the group. The function \( \Delta(g) = \det Ad(g) \) is called the modular function of \( G \). For a unimodular Lie group, \( \Delta(g) = 1 \).

2.2.2 Integration and Differentiation on Unimodular Lie Groups

Unimodular Lie groups are defined by the fact that their integration measures are invariant under shifts and inversions. In any parametrization, this measure (or the corresponding volume element) can be expressed as in the examples by first computing a left or right Jacobian matrix and then setting \( dg = |J(q)|dq_1 dq_2 \cdots dq_n \) where \( n \) is the dimension of the group. In the special case when \( q = x \) is the exponential coordinates, then \( 37 \)

\[
\int_G f(q) dq = \int_\mathcal{G} f(e^X) \det \left( \frac{1 - e^{-ad(X)}}{ad(X)} \right) dx
\]

where \( x = X^\vee \) and \( dx = dx_1 dx_2 \cdots dx_n \). In the above expression it makes sense to write the division of one matrix by another because the involved matrices commute. The symbol \( \mathcal{G} \) is used to denote the Lie algebra corresponding to \( G \). In practice the integral is performed over a subset of \( \mathcal{G} \), which is equivalent to defining \( f(e^X) \) to be zero over some portion of \( \mathcal{G} \).

Let \( f(g) \) be a probability density function (or pdf for short) on a Lie group \( G \). Then

\[
\int_G f(g) dg = 1 \quad \text{and} \quad f(g) \geq 0.
\]
It can be shown that unimodularity implies the following equalities for arbitrary \( h \in G \), which generally do not all hold simultaneously for measures on nonunimodular Lie groups:

\[
\int_G f(g^{-1})dg = \int_G f(h \circ g)dg = \int_G f(g \circ h)dg = \int_G f(g)dg. \tag{21}
\]

Many different kinds of unimodular Lie groups exist. For example, \( SO(3) \) is compact and therefore has finite volume; \( SE(2) \) belongs to a class of Lie groups that are called solvable, \( H(1) \) belongs to a class called nilpotent; and \( SL(2, \mathbb{R}) \) belongs to a class called semisimple. Each of these classes of Lie groups has been studied extensively. But for the purpose of this discussion, it is sufficient treat them all within the larger class of unimodular Lie groups.

Given a function \( f(g) \), the left and right Lie derivatives are defined with respect to any basis element of the Lie algebra \( X_i \in \mathcal{G} \) as

\[
\begin{align*}
\hat{X}_i^r f(g) &= \left( \frac{d}{dt} f(g \circ \exp(tX_i)) \right) \bigg|_{t=0} \quad \text{and} \quad 
\hat{X}_i^l f(g) &= \left( \frac{d}{dt} f(\exp(-tX_i) \circ g) \right) \bigg|_{t=0}. 
\end{align*} \tag{22}
\]

The use of \( l \) and \( r \) mimicks the way that the subscripts were used in the Jacobians \( J_l \) and \( J_r \), in the sense that if \( \exp(tX_i) \) appears on the left/right then the corresponding derivative is given an \( l/r \) designation. This notation, while not standard in the mathematics literature, is useful in computations because when evaluating left/right Lie derivatives in coordinates \( g = g(q) \), the left/right Jacobians enter in the computation as:

\[
\hat{X}_r f = [J_r(q)]^{-T} \nabla_q f \quad \text{and} \quad \hat{X}_l f = -[J_l(q)]^{-T} \nabla_q f \tag{23}
\]

where \( \hat{X} = [\hat{X}_1, ..., \hat{X}_n]^T \), \( \hat{X}_l = [\hat{X}_1, ..., \hat{X}_n]^T \), and \( \nabla_q = [\frac{\partial}{\partial q_1}, ..., \frac{\partial}{\partial q_n}]^T \) is the gradient operator treating \( q \) like Cartesian coordinates.

### 2.3 Probability Theory and Harmonic Analysis on Unimodular Lie Groups

Given two probability density functions \( f_1(g) \) and \( f_2(g) \), their convolution is

\[
(f_1 * f_2)(g) = \int_G f_1(h) f_2(h^{-1} \circ g) dh. \tag{24}
\]

Here \( h \in G \) is a dummy variable of integration. Convolution inherits associativity from the group operation, but since in general \( g_1 \circ g_2 \neq g_2 \circ g_1 \), \( (f_1 * f_2)(g) \neq (f_2 * f_1)(g) \).

For a unimodular Lie group, the convolution integral of the form in \( [24] \) can be written in the following equivalent ways:

\[
(f_1 * f_2)(g) = \int_G f_1(z^{-1}) f_2(z \circ g) dz
= \int_G f_1(g \circ k^{-1}) f_2(k) dk \tag{25}
\]

where the substitutions \( z = h^{-1} \) and \( k = h^{-1} \circ g \) have been made, and the invariance of integration under shifts and inversions in \( [21] \) is used.

A powerful generalization of classical Fourier analysis exists. It is built on families of unitary matrix-valued functions of group-valued argument that are parametrized by values \( \lambda \) drawn from a set \( \hat{G} \) and satisfy the homomorphism property:

\[
U(g_1 \circ g_2, \lambda) = U(g_1, \lambda) U(g_2, \lambda). \tag{26}
\]
Using \( \ast \) to denote the Hermitian conjugate, it follows that
\[
I = U(e, \lambda) = U(g^{-1} \circ g, \lambda) = U(g^{-1}, \lambda)U(g, \lambda),
\]
and so
\[
U(g^{-1}, \lambda) = (U(g, \lambda))^{-1} = U^*(g, \lambda).
\]

In this generalized Fourier analysis (called noncommutative harmonic analysis) each \( U(g, \lambda) \) is constructed to be \textit{irreducible} in the sense that it is not possible to simultaneously block-diagonalize \( U(g, \lambda) \) by the same similarity transformation for all values of \( g \) in the group. Such a matrix function \( U(g, \lambda) \) is called an \textit{irreducible unitary representation}. Completeness of a set of representations means that every (reducible) representation can be decomposed into a direct sum of the representations in the set.

Once a complete set of IURs is known for a unimodular Lie group, the Fourier transform of a function on that group can be defined as
\[
\hat{f}(\lambda) = \int_G f(g)U(g^{-1}, \lambda)dg.
\]

Here \( \lambda \) (which can be thought of as frequency) indexes the complete set of all IURs. An inversion formula can be used to recover the original function from all of the Fourier transforms as
\[
f(g) = \int_G \text{trace}[\hat{f}(\lambda)U(g, \lambda)]d(\lambda).
\]

The integration measure \( d(\lambda) \) on the dual (frequency) space \( \hat{G} \) is very different from one group to another. In the case of a compact Lie group, \( \hat{G} \) is discrete, and the resulting inversion formula is a series, much like the classical Fourier series for \( 2\pi \)-periodic functions.

A convolution theorem follows from (26) as
\[
(\hat{f}_1 \ast \hat{f}_2)(\lambda) = \hat{f}_2(\lambda)\hat{f}_1(\lambda)
\]
and so does the Parseval/Plancherel formula:
\[
\int_G |f(g)|^2dg = \int_G ||\hat{f}(\lambda)||^2d(\lambda).
\]

Here \( || \cdot || \) is the Hilbert-Schmidt (Frobenius) norm, and \( d(\lambda) \) is the dimension of the matrix \( U(g, \lambda) \).

A useful definition is
\[
u(X_i, \lambda) = \frac{d}{dt} (U(\exp(tX_i), \lambda))|_{t=0}.
\]

Explicit expressions for \( U(g, \lambda) \) and \( u(X_i, \lambda) \) using the exponential map and corresponding parameterizations for the groups \( SO(3), SE(2) \) and \( SE(3) \) are given in [58, 32].

As a consequence of these definitions, it can be shown that the following operational properties result [24]:
\[
\hat{X}_i f = u(X_i, \lambda)\hat{f}(\lambda) \quad \text{and} \quad \hat{X}_i f = -\hat{f}(\lambda)u(X_i, \lambda).
\]

This is very useful in probability problems because a diffusion equation with drift of the form
\[
\frac{\partial \rho(g; t)}{\partial t} = -\sum_{i=1}^d h_i(t) \hat{X}_i \rho(g; t) + \frac{1}{2} \sum_{i,j=1}^d D_{ij} \hat{X}_i \hat{X}_j \rho(g; t)
\]

(29)
(where $D = [D_{ij}]$ is symmetric and positive semidefinite and given initial conditions $\rho(g; 0) = \delta(g)$) can be solved in the dual space $\hat{G}$, and then the inversion formula can convert it back. Explicitly,

$$\rho(g; t) = \int_{\hat{G}} \text{trace}[\exp(t{\mathcal{B}(\lambda)}))]U(g, \lambda)]d(\lambda)$$

(30)

where

$${\mathcal{B}(\lambda)} = \frac{1}{2} \sum_{k,l=1}^{n} D_{lk} u(X_l, \lambda)u(X_k, \lambda) - \sum_{l=1}^{n} h_l u(X_l, \lambda).$$

The solution to this sort of diffusion equation is important as a generalization of the concept of a Gaussian distribution. It has been studied extensively in the case of $G = SE(3)$ in the context of polymer statistical mechanics and robotic manipulators [22, 23, 82]. As will be shown shortly, some of the classical information-theoretic inequalities that follow from the Gaussian distribution can be computed using the above analysis.

3 Properties of Entropy and Relative Entropy on Groups

As defined earlier, the entropy of a pdf on a unimodular Lie group is

$$S(f) = - \int_{G} f(g) \log f(g) \, dg.$$  

For example, the entropy of a Gaussian distribution with covariance $\Sigma$ is

$$S(\rho(g; t)) = \log\{(2\pi e)^{n/2}|\Sigma(t)|^{\frac{1}{2}}\}$$  

(31)

where $\log = \log_e$.

The Kullback-Leibler distance between the pdfs $f_1(g)$ and $f_2(g)$ on a Lie group $G$ naturally generalizes from its form in $\mathbb{R}^n$ as

$$D_{KL}(f_1 \parallel f_2) = \int_{G} f_1(g) \log \left(\frac{f_1(g)}{f_2(g)}\right) \, dg.$$  

(32)

As with the case of pdfs in $\mathbb{R}^n$, $D_{KL}(f_1 \parallel f_2) \geq 0$ with equality when $D_{KL}(f \parallel f) = 0$. And if $D_{KL}(f_1 \parallel f_2) = 0$ then $f_1(g) = f_2(g)$ at “almost all” values of $g \in G$ (or, in probability terminology “$f_1(g) = f_2(g)$ almost surely”). That is, they must be the same up to a set of measure zero.

Something that is not true in $\mathbb{R}^n$ that holds for a compact Lie group is that the limiting distribution is the number one. If $f_2(g) = 1$ is the limiting distribution, then $D_{KL}(f_1 \parallel 1) = -S(f_1)$.

3.1 Convolutions Generally Increase Entropy

**Theorem 3.1**: Given pdfs $f_1(g)$ and $f_2(g)$ on the unimodular Lie group $G$,

$$S(f_1 * f_2) \geq \max\{S(f_1), S(f_2)\}.$$  

(33)
**Proof:** Denote the result of an $n$-fold convolution on $G$ as

$$f_{1,n}(g) = (f_1 * f_2 * f_3 * \cdots * f_n)(g).$$

Recall that a single pairwise convolution is computed as

$$f_{i,i+1}(g) = (f_i * f_{i+1})(g) = \int_G f_i(h) f_{i+1}(h^{-1} g) dh = \int_G f_i(g k^{-1}) f_{i+1}(k) dk \neq (f_{i+1} * f_i)(g).$$

The $n$-fold convolution can be computed by performing a series of pairwise convolutions and stringing them together using the associative law. Convolution of functions on the group inherits associativity from the group law, which is reflected in the notation

$$f_{i,i+2} = (f_i * f_{i+1} * f_{i+2})(g) = (f_i * f_{i+1,i+2})(g) = (f_{i,i+1} * f_{i+2})(g)$$

where

$$(f_i f_{i+1,i+2})(g) = (f_i(f_{i+1} f_{i+2}))(g) \quad \text{and} \quad (f_{i,i+1} f_{i+2})(g) = ((f_i f_{i+1}) f_{i+2})(g).$$

Johnson and Suhov [40, 41] proved the following result for compact Lie groups:

$$D_{KL}(f_{1,n} \| 1) - D_{KL}(f_{1,n-1} \| 1) = - \int_G D_{KL}(f_{1,n-1} \| R(h) f_{1,n}) f_n(h) dh$$

(34)

where $(R(h)f)(g) = f(g \circ h)$ is the right shift operator. Since the integrand on the right side of (34) is nonnegative at all values of $h$ (and in fact, strictly positive unless all $f_i(g)$ are delta functions), this indicates that

$$D_{KL}(f_{1,n} \| 1) \leq D_{KL}(f_{1,n-1} \| 1)$$

with equality only holding in pathological cases. And so iterated convolutions lead to

$$\lim_{n \to \infty} D_{KL}(f_{1,n} \| 1) = 0 \implies f_{1,n}(g) = 1 \text{ a.s.}$$

A noncompact group can not have $f(g) = 1$ as a limiting distribution, and so it does not make sense in this case to use the notation $D_{KL}(f_{1,n} \| 1)$. Nevertheless, essentially the same proof that gives (34) can be used in the more general case of not-necessarily-compact unimodular Lie groups to show that entropy must increase as a result of convolution. This can be observed by first expanding out $S(f_{1,n})$ as:

$$S(f_{1,n}) = - \int_G f_{1,n}(g) \log f_{1,n}(g) dg$$

(35)

$$= - \int_G (f_{1,n-1} * f_n)(g) \log f_{1,n}(g) dg$$

(36)

$$= - \int_G \left[ \int_G f_{1,n-1}(g \cdot h^{-1}) f_n(h) dh \right] \log f_{1,n}(g) dg$$

(37)

$$= - \int_G \int_G f_{1,n-1}(g \cdot h^{-1}) f_n(h) \log f_{1,n}(g) dg dh$$

(38)

$$= - \int_G \int_G f_{1,n-1}(k) f_n(h) \log f_{1,n}(k \cdot h) dk dh.$$  

(39)
In going from (37) to (38) all that was done was to reverse the order of integration (i.e., using Fubini’s Theorem) and in going from (38) to (39) the change of variables \( k = g \circ h^{-1} \) is used together with the invariance of integration under shifts.

Next, observe that

\[
S(f_{1,n-1}) = -\int_{G} f_{1,n-1}(k) \log f_{1,n-1}(k) dk
\]

\[
= -\left( \int_{G} f_{1,n-1}(k) \log f_{1,n-1}(k) dk \right) \left( \int_{G} f_n(h) dh \right)
\]

\[
= -\int_{G} \left( \int_{G} f_{1,n-1}(k) \log f_{1,n-1}(k) dk \right) f_n(h) dh.
\]

and so

\[
S(f_{1,n}) - S(f_{1,n-1}) = \int_{G} \left( \int_{G} f_{1,n-1}(k) [\log f_{1,n-1}(k) - \log f_{1,n}(k \circ h)] dk \right) f_n(h) dh
\]

\[
= \int_{G} \left( \int_{G} f_{1,n-1}(k) \log \frac{f_{1,n-1}(k)}{f_{1,n}(k \circ h)} dk \right) f_n(h) dh
\]

\[
= \int_{G} D_{KL}(f_{1,n-1} \| R(h)f_{1,n}) f_n(h) dh
\]

\[
\geq 0.
\]

Since no direct comparison between \( f_{1,n} \) and the uniform distribution is made, Johnson and Suhov’s proof of (34) that has been adapted above yields

\[
S(f_{1,n-1} * f_n) \geq S(f_{1,n-1}).
\]

Essentially the same proof can be used to show that

\[
S(f_1 * f_{2,n}) \geq S(f_{2,n}).
\]

In other words, convolution in either order increases entropy.

### 3.2 Entropy Inequalities from Jensen’s Inequality

Jensen’s inequality is a fundamental tool that is often used in deriving information-theoretic inequalities, as well as inequalities in the field of convex geometry. In the context of Lie groups, Jensen’s inequality can be written as

\[
\Phi \left( \int_{G} \phi(g) \rho(g) dg \right) \leq \int_{G} \Phi(\phi(g)) \rho(g) dg
\]

where \( \Phi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R} \) is a convex function on the half infinite line, \( \rho(g) \) is a pdf, and \( \phi(g) \) is another nonnegative measurable function on \( G \).

Two important examples of \( \Phi(x) \) are \( \Phi_1(x) = -\log x \) and \( \Phi_2(x) = +x \log x \). If \( G \) is compact, any constant function on \( G \) is measurable. Letting \( \phi(g) = 1 \) and \( \Phi(x) = \Phi_2(x) \) then gives \( 0 \leq -S(f) \) for a pdf \( f(g) \). In contrast, for any unimodular Lie group, letting \( \rho(g) = f(g) \), \( \phi(g) = [f(g)]^\alpha \) and \( \Phi(x) = \Phi_1(x) \) gives

\[
- \log \left( \int_{G} [f(g)]^{1+\alpha} dg \right) \leq \alpha S(f).
\]
This leads to the following theorem.

**Theorem 3.2:** Let \( \| \hat{f}(\lambda) \| \) denote the Frobenius norm and \( \| \hat{f}(\lambda) \|_2 \) denote the induced 2-norm of the Fourier transform of \( f(g) \) and define

\[
D_2(f) = - \int_G \log \| \hat{f}(\lambda) \|_2^2 d(\lambda), \quad D(f) = - \int_G \log \| \hat{f}(\lambda) \|^2 d(\lambda), \quad \tilde{D}(f) = - \log \int_G \| \hat{f}(\lambda) \|^2 d(\lambda). \tag{42}
\]

Then

\[
S(f) \geq \tilde{D}(f) \quad \text{and} \quad D(f) \leq D_2(f) \tag{43}
\]

and

\[
D_2(f_1 * f_2) \geq D_2(f_1) + D_2(f_2) \quad \text{and} \quad D(f_1 * f_2) \geq D(f_1) + D(f_2). \tag{44}
\]

Furthermore, denote the unit Heaviside step function on the real line as \( u(x) \) and let

\[
B = \int_G u \left( \| \hat{f}(\lambda) \| \right) d(\lambda). \quad \text{Then} \quad \tilde{D}(f) + \log B \leq D(f)/B. \tag{45}
\]

For finite groups \( B = 1 \) for functions that have full spectrum, and for bandlimited expansions on other groups \( B \) is finite.

**Proof:** Substituting \( \alpha = 1 \) into (41) and using the Plancherel formula (28) yields

\[
S(f) \geq - \log \left( \int_G |f(g)|^2 dg \right) = - \log \left( \int_G \| \hat{f}(\lambda) \|^2 d(\lambda) \right) = \tilde{D}(f).
\]

The fact that \(- \log x\) is a decreasing function and \( \|A\|_2 \leq \|A\| \) for all \( A \in \mathbb{C}^{n \times n} \) gives the second inequality in (43).

The convolution theorem together with the facts that both norms are submultiplicative, \(- \log(x)\) is a decreasing function, and the log of the product is the sum of the logs gives

\[
D(f_1 * f_2) = - \int_G \log \| \hat{f}_1 * \hat{f}_2(\lambda) \|^2 d(\lambda) = - \int_G \log \| \hat{f}_1(\lambda) \hat{f}_2(\lambda) \|^2 d(\lambda) \geq D(f_1) + D(f_2).
\]

An identical calculation follows for \( D_2 \). The statement in (45) follows from the Plancherel formula (28) and using Jensen’s inequality (40) in the dual space \( \hat{G} \) rather than on \( G \):

\[
\Phi \left( \int_{\hat{G}} \| \hat{\phi}(\lambda) \| \rho(\lambda) d(\lambda) \right) \leq \int_{\hat{G}} \Phi(\| \hat{\phi}(\lambda) \|) \rho(\lambda) d(\lambda) \quad \text{where} \quad \int_{\hat{G}} \rho(\lambda) d(\lambda) = 1 \quad \text{and} \quad \rho(\lambda) \geq 0. \tag{46}
\]

Recognizing that when \( B \) is finite \( \rho(\lambda) = u \left( \| \hat{f}(\lambda) \| \right) / B \) becomes a probability measure on this dual space, it follows that

\[
\tilde{D}(f) = - \log \left( \int_{\hat{G}} \| \hat{f}(\lambda) \|^2 d(\lambda) \right) = - \log \left( B \int_{\hat{G}} \| \hat{f}(\lambda) \|^2 \rho(\lambda) d(\lambda) \right)
\]

\[
\leq - \log B - \int_{\hat{G}} \log \left( \| \hat{f}(\lambda) \|^2 \right) \rho(\lambda) d(\lambda) = - \log B + D(f)/B.
\]

This completes the proof.

Properties of dispersion measures similar to \( D(f) \) and \( D_2(f) \) were studied in [35], but no connections to entropy were provided previously. By definition, bandlimited
expansions have $B$ finite. On the other hand, it is a classical result that for a finite group, $\Gamma$, the Plancherel formula is (see, for example, [24]):

$$\sum_{\gamma \in \Gamma} |f(\gamma)|^2 = \frac{1}{|\Gamma|} \sum_k d_k^2 \|\hat{f}_k\|^2$$

where $\alpha$ is the number of conjugacy classes of $\Gamma$ and $d_k$ is the dimension of $\hat{f}_k$. And by Burnside’s formula $\sum_{k=1}^\alpha d_k^2 = |\Gamma|$ it follows that $B = 1$ when all $\|\hat{f}_k\| \neq 0$.

### 3.3 The Entropy Produced by Convolution on a Finite Group is Bounded

Let $\Gamma$ be a finite group with $|\Gamma|$ elements $\{g_1, ..., g_{|\Gamma|}\}$, and let $\rho^\Gamma(g_i) \geq 0$ with $\sum_{i=1}^{|\Gamma|} \rho^\Gamma(g_i) = 1$ define a probability density/distribution on $\Gamma$. In analogy with how convolution and entropy are defined on a Lie group, $G$, they can also be defined on a finite group, $\Gamma$ by using the Dirac delta function for $G$, denoted here as $\delta(g)$. If $\Gamma < G$ (i.e., if $\Gamma$ is a subgroup of $G$), then letting

$$\rho^G = \sum_{i=1}^{|\Gamma|} \rho^\Gamma(g_i) \delta(g_i^{-1} \circ g) = \sum_{\gamma \in \Gamma} \rho^\Gamma(\gamma) \delta(\gamma^{-1} \circ g)$$

can be used to define a pdf on $G$ that is equivalent to a pdf on $\Gamma$ in the sense that if the convolution of two pdfs on $\Gamma$ is

$$(\rho^\Gamma_1 * \rho^\Gamma_2)(g_i) = \sum_{j=1}^{|\Gamma|} \rho^\Gamma_1(g_j) \rho^\Gamma_2(g_j^{-1} \circ g_i) \quad (47)$$

then

$$(\rho^G_1 * \rho^G_2)(g) = \sum_{\gamma \in \Gamma} (\rho^\Gamma_1 * \rho^\Gamma_2)(\gamma) \delta(\gamma^{-1} \circ g). \quad (48)$$

Given a finite group, $\Gamma$, let

$$S(\rho) = -\sum_{i=1}^{|\Gamma|} \rho(g_i) \log \rho(g_i) = -\sum_{\gamma \in \Gamma} \rho(\gamma) \log \rho(\gamma).$$

Unlike the case of differential/continuous entropy on a Lie group, $0 \leq S(\rho)$.

The following theorem describes how the discrete entropy of pdfs on $\Gamma$ behaves under convolution. Since only finite groups are addressed, the superscript $\Gamma$ on the discrete values $\rho(g_i)$ are dropped.

**Theorem 3.3:** The entropy of the convolution of two pdfs on a finite group is greater than either of the entropies of the convolved pdfs and is no greater than the sum of their individual entropies

$$\max\{S(\rho_1), S(\rho_2)\} \leq S(\rho_1 * \rho_2) \leq S(\rho_1) + S(\rho_2). \quad (49)$$

**Proof:** The lower bound follows in the same way as the proof given for Theorem 1.1 with summation in place of integration. The entropy of convolved distributions on a finite group can be bounded from above in the following way.
Since the convolution sum contains products of all pairs, and each product is positive, it follows that
\[
\rho_1(g_k)\rho_2(g^{-1}_k \circ g_i) \leq (\rho_1 \ast \rho_2)(g_i)
\]
for all \(k \in \{1, \ldots, |\Gamma|\}\). Therefore, since \(\log\) is a strictly increasing function, it follows that
\[
-S(\rho_1 \ast \rho_2) \geq \sum_{i=1}^{|\Gamma|} \left( \sum_{j=1}^{|\Gamma|} \rho_1(g_j)\rho_2(g^{-1}_j \circ g_i) \right) \log \left( \rho_1(g_k)\rho_2(g^{-1}_k \circ g_i) \right).
\]
Since this is true for all values of \(k\), we can bring the \(\log\) term inside of the summation and choose \(k = j\). Then multiplying by \(-1\), and using the properties of the \(\log\) function, we get
\[
S(\rho_1 \ast \rho_2) \leq -\sum_{i=1}^{|\Gamma|} \sum_{j=1}^{|\Gamma|} \rho_1(g_j)\rho_2(g^{-1}_j \circ g_i) \log \rho_1(g_j) - \sum_{i=1}^{|\Gamma|} \sum_{j=1}^{|\Gamma|} \rho_1(g_j)\rho_2(g^{-1}_j \circ g_i) \log \rho_2(g^{-1}_j \circ g_i).
\]
Rearranging the order of summation signs gives
\[
S(\rho_1 \ast \rho_2) \leq -\sum_{j=1}^{|\Gamma|} \rho_1(g_j) \log \rho_1(g_j) \left( \sum_{i=1}^{|\Gamma|} \rho_2(g^{-1}_j \circ g_i) \right) - \sum_{j=1}^{|\Gamma|} \rho_1(g_j) \left( \sum_{i=1}^{|\Gamma|} \rho_2(g^{-1}_j \circ g_i) \right) \log \rho_2(g^{-1}_j \circ g_i).
\]
But summation of a function over a group is invariant under shifts. That is,
\[
\sum_{i=1}^{|\Gamma|} F(g^{-1}_j \circ g_i) = \sum_{i=1}^{|\Gamma|} F(g_i) \quad \text{or} \quad \sum_{\gamma \in \Gamma} F(\gamma^{-1} \circ g) = \sum_{\gamma \in \Gamma} F(g).
\]
Hence, the terms in parenthesis in (50) can be written by replacing \(g^{-1}_j \circ g_i\) with \(g_i\) gives (50).

3.4 Entropy and Decompositions

Aside from the ability to sustain the concept of convolution, one of the fundamental ways that groups resemble Euclidean space is the way in which they can be decomposed. In analogy with the way that an integral over a vector-valued function with argument \(x \in \mathbb{R}^n\) can be decomposed into integrals over each coordinate, integrals over Lie groups can also be decomposed in natural ways. This has implications with regard to inequalities involving the entropy of pdfs on Lie groups. Analogous expressions hold for finite groups, with volume replaced by the number of group elements.

3.4.1 Direct Products

Given the direct product of two groups, \(G_1 \times G_2\), and a probability density \(f(g_1, g_2)\) with
\[
\int_G \int_G f(g_1, g_2) dg_1 dg_2 = 1
\]
and the corresponding entropy is
\[
S_{12} = -\int_G \int_G f(g_1, g_2) \log f(g_1, g_2) dg_1 dg_2.
\]
Inexact analogy with classical information theory, we can write:

$$S_{12} \leq S_1 + S_2 \quad (51)$$

where

$$f_1(g_1) = \int_G f(g_1, g_2) dg_2 \quad \text{and} \quad f_2(g_2) = \int_G f(g_1, g_2) dg_1,$$

and

$$S_i = -\int_G f_i(g_i) \log f_i(g_i) dg_i.$$  

Equality in (51) holds if and only if

$$f(g_1, g_2) = f_1(g_1) f_2(g_2).$$

As in the case of pdfs on Euclidean space, (51) follows from the fact that the Kullback-Leibler divergence in (32) has the property that

$$D_{KL}(f \parallel f_1 f_2) \geq 0.$$  

3.4.2 Coset Decompositions

Given a subgroup $H \leq G$, and any element $g \in G$, the left coset $gH$ is defined as $gH = \{ g \circ h | h \in H \}$. Similarly, the right coset $Hg$ is defined as $Hg = \{ h \circ g | h \in H \}$. In the special case when $g \in H$, the corresponding left and right cosets are equal to $H$. More generally for all $g \in G$, $g \in gH$ and $g_1 H = g_2 H$ if and only if $g_2^{-1} \circ g_1 \in H$. Likewise for right cosets $Hg_1 = Hg_2$ if and only if $g_1 \circ g_2^{-1} \in H$. Any group is divided into disjoint left (right) cosets, and the statement “$g_1$ and $g_2$ are in the same left (right) coset” is an equivalence relation. An important property of $gH$ and $Hg$ is that they have the same number of elements as $H$. Since the group is divided into disjoint cosets, each with the same number of elements, it follows that the number of cosets must divide without remainder the number of elements in the group. The set of all left (or right) cosets is called the left (or right) coset space, and is denoted as $G/H$ (or $H \backslash G$). For finite groups one writes $|G/H| = |H \backslash G| = |G|/|H|$. This result is called Lagrange’s theorem. Similar expressions can be written for Lie groups and Lie subgroups after the appropriate concept of volume is introduced. We will use the following well-known fact [37]:

$$\int_G f(g) dg = \int_{G/H} \left( \int_H f(g \circ h) dh \right) d(gH) \quad (52)$$

where $g \in gH$ is taken to be the coset representative. In the special case when $f(g)$ is a left-coset function (i.e., a function that is constant on left cosets), (52) reduces to

$$\int_G f(g) dg = \int_{G/H} F(gH) d(gH)$$

where it is assumed that $d(h)$ is normalized so that $\text{Vol}(H) = \int_H dh = 1$, and

$$F(gH) = \int_H f(g \circ h) dh$$

is the value of the function $f(g)$ on each coset representative (which is the same as that which results from averaging over the coset $gH$).

**Theorem 3.4**: The entropy of a pdf on a unimodular Lie group is no greater than the sum of the marginal entropies on a subgroup and the corresponding coset space:

$$S(f_G) \leq S(f_{G/H}) + S(f_H). \quad (53)$$
Proof: For the moment it will be convenient to denote a function on $G$ as $f_G(g)$ (rather than $f(g)$) and write

$$f_G(g) = f_{G/H \times H}(g) = \tilde{f}_{G/H \times H}(gH, e).$$

That is, a function on $G$ evaluated at $g$ can be equally described as a function on a coset, together with a rule for extracting a specific coset representative, which in this case is the identity. This means that given $gH$, $g$ is recovered from $g \in gH$ as $g \circ e^{-1} = g$. By enforcing the constraint on the definition of $\tilde{f}_{G/H \times H}$ that

$$f_G(g \circ h) = \tilde{f}_{G/H \times H}(gH, h) \quad \text{and} \quad \tilde{f}_{G/H \times H}(H, h) = \tilde{f}_{G/H \times H}(H, e),$$

then $g$ can be recovered from $g \circ h \in gH$ as $g \circ h \circ h^{-1} = g$. Using this construction, we can define

$$f_H(h) = \int_{G/H} f_G(g \circ h) d(gH) = \int_{G/H} \tilde{f}_{G/H \times H}(gH, h) d(gH)$$

and

$$\tilde{f}_{G/H}(gH) = \int_H f_G(g \circ h) dh = \int_H \tilde{f}_{G/H \times H}(gH, h) dh.$$

For example, if $G = SE(n)$ is a Euclidean motion group and $H = SO(n)$ is the subgroup of pure rotations in $n$-dimensional Euclidean space, then $G/H \cong \mathbb{R}^n$, and we can write

$$\int_{SE(n)} f(g) d(g) = \int_{SE(n)/SO(n)} \left( \int_{SO(n)} f(g \circ h) d(R) \right) d(t)$$

It follows from the classical information-inequality for the entropy of marginal distributions obtained by letting $F(g) = -f(g) \log f(g)$ and using the nonnegativity of the Kullback-Leibler divergence

$$D(f_G(g \circ h) \| f_{G/H \cdot f_H(h)}) \geq 0$$

together with the shift-invariance of integrals on unimodular Lie groups that (53) holds.

3.4.3 Double Coset Decompositions

Let $H < G$ and $K < G$. Then for any $g \in G$, the set

$$HgK = \{ h \circ g \circ k | h \in H, k \in K \}$$

(54)

is called the double coset of $H$ and $K$, and any $g' \in HgK$ (including $g' = g$) is called a representative of the double coset. Though a double coset representative often can be described with two or more different pairs $(h_1, k_1)$ and $(h_2, k_2)$ so that $g' = h_1 \circ g \circ k_1 = h_2 \circ g \circ k_2$, we only count $g$ once in $HgK$. Hence $|HgK| \leq |G|$, and in general $|HgK| \neq |H| \cdot |K|$. In general, the set of all double cosets of $H$ and $K$ is denoted $H/G \setminus K$. Hence we have the hierarchy $g \in HgK \in H \setminus G \setminus K$. It can be shown that membership in a double coset is an equivalence relation. That is, $G$ is partitioned into disjoint double cosets, and for $H < G$ and $K < G$ either $Hg_1K \cap Hg_2K = \emptyset$ or $Hg_1K = Hg_2K$. 

21
Another interesting thing to note (when certain conditions are met) is the decom-
position of the integral of a function on a group in terms of two subgroups and a double
coset space:

\[
\int_{G} F(g) d(g) = \int_{K} \int_{K \setminus G/H} \int_{H} F(k \circ g \circ h) d(h) d(KgH) d(k).
\] (55)

A particular example of this is the integral over SO(3), which can be written in
terms of Euler angles as

\[
\int_{SO(3)} dg = \int_{0}^{2\pi} \int_{0}^{\pi} \frac{1}{8\pi^2} \sin\beta d\alpha d\beta d\gamma =
\int_{SO(2)} \int_{SO(2) \setminus SO(3) / SO(2)} \int_{SO(2)} \left( \frac{1}{2\pi^2} d\alpha \right) \left( \frac{1}{2} \sin\beta d\beta \right) \left( \frac{1}{2\pi} d\gamma \right).
\]

**Theorem 3.5:** The entropy of a pdf on a group is no greater than the sum of marginal
entropies over any two subgroups and the corresponding double-coset space:

\[
S(f_G) \leq S(f_K) + S(f_{K \setminus G/H}) + S(f_H).
\] (56)

**Proof:** Consistent with (55) it is possible to decompose a function \( f_G(g) \)
as

\[ f_G(g) = \int_{K \setminus K \setminus G/H \times H} f_{K \setminus K \setminus G/H \times H}(e, KgH, e) \] where \( f_G(k \circ g \circ h) = \int_{K \setminus K \setminus G/H \times H} f_{K \setminus K \setminus G/H \times H}(k, KgH, h). \)

If

\[
\begin{align*}
\int_{K \setminus G/H} f_K(k) d(k) &= \int_{K \setminus G/H} \int_{H} f_G(k \circ g \circ h) d(h) d(KgH) \\
\int_{K \setminus G/H} f_H(h) d(h) &= \int_{K \setminus G/H} \int_{K \setminus G} f_G(k \circ g \circ h) d(k) d(KgH)
\end{align*}
\]

and

\[
\begin{align*}
\int_{K \setminus G/H} f_{K \setminus G/H} &= \int_{K \setminus G/H} \int_{H} f_G(k \circ g \circ h) d(h) d(k)
\end{align*}
\]

then letting \( F(g) = -f(g) \log f(g) \) and using the nonnegativity of the Kullback-Leibler diver-
gence

\[ D(f_G(k \circ g \circ h) \parallel f_K(k) \cdot f_{K \setminus G/H} \cdot f_H(h)) \geq 0 \]

together with the shift-invariance of integrals on unimodular Lie groups gives (56).

### 3.4.4 Nested Coset Decompositions

**Theorem 3.6:** The entropy of a pdf is no greater than the sum of entropies of its marginals
over coset spaces defined by nested subgroups:

\[
S(f_G) \leq S(f_{G/K}) + S(f_{K/H}) + S(f_H).
\] (57)

**Proof:** Given a subgroup \( K \) of \( H \), which is itself a subgroup of \( G \) (that is, \( H < K < G \)), it is possible to write (57)

\[
\begin{align*}
\int_{G/H} F(gH) d(gH) &= \int_{G/K} \int_{K/H} F(g \circ kH) d(kH) d(gK).
\end{align*}
\]

Therefore,

\[
\begin{align*}
\int_{G} F(g) d(g) &= \int_{G/K} \int_{K/H} \int_{H} F(g \circ k \circ h) d(h) d(kH) d(gK).
\end{align*}
\]
Again letting $F(g) = -f_G(g) \log f_G(g)$, it follows from the properties of Kullback-Leibler divergence and the unimodularity of $G$ that if

$$f_{G/K}(gK) = \int_{K/H} \int_{H} f(g \circ k \circ h) dh dk$$

and

$$f_{K/H}(kH) = \int_{G/K} \int_{H} f(g \circ k \circ h) dh dk$$

then (57) follows.

### 3.4.5 Class Functions and Normal Subgroups

In analogy with the way a coset is defined, the conjugate of a subgroup $H$ for a given $g \in G$ is defined as $gHg^{-1} = \{g \circ h \circ g^{-1} | h \in H\}$. Recall that a subgroup $N \leq G$ is called normal if and only if $gNg^{-1} \subseteq N$ for all $g \in G$. This is equivalent to the conditions $g^{-1}Ng \subseteq N$, and so we also write $gNg^{-1} = N$ and $gN = Ng$ for all $g \in G$.

A function, $\chi(g)$, that is constant on each class has the property that

$$\chi(g) = \chi(h^{-1} \circ g \circ h) \quad \text{or} \quad \chi(h \circ g) = \chi(g \circ h)$$

(58)

for any $g, h \in G$. Though convolution of functions on a noncommutative group is generally noncommutative, the special nature of class functions means that

$$(f \ast \chi)(g) = \int_{G} f(h) \chi(h^{-1} \circ g) dh = \int_{G} f(h) \chi(g \circ h^{-1}) dh$$

$$= \int_{G} \chi(k) f(k^{-1} \circ g) dk = (\chi \ast f)(g).$$

where the change of variables $k = g \circ h^{-1}$ is used together with the unimodularity of $G$.

### 3.5 When Inequivalent Convolutions Produce Equal Entropy

In general $(\rho_1 \ast \rho_2)(g) \neq (\rho_2 \ast \rho_1)(g)$. Even so, it can be the case that $S(\rho_1 \ast \rho_2)(g) = S(\rho_2 \ast \rho_1)(g)$. This section addresses several special cases when this equality holds.

Let $G$ denote a unimodular Lie group and for arbitrary $g, g_1 \in G$ define $\rho_g'(g) = \rho(g^{-1})$, $L_{g_1} \rho(g) = \rho(g_1^{-1} \circ g)$, $R_{g_1} \rho(g) = \rho(g \circ g_1)$, $C_{g_1} \rho(g) = \rho(g_1^{-1} \circ g \circ g_1)$. Then if $\rho(g)$ is a pdf, it follows immediately from (21) that $\rho_g'(g)$, $L_{g_1} \rho(g)$, $R_{g_1} \rho(g)$, and $C_{g_1} \rho(g)$ are all pdfs. A function for which $\rho_g'(g) = \rho(g)$ is called symmetric, whereas a function for which $C_{g_1} \rho(g) = \rho(g)$ for all $g_1 \in G$ is a class function (i.e., it is constant on conjugacy classes).

**Theorem 3.7:** For arbitrary pdfs on a unimodular Lie group $G$ and arbitrary $g_1, g_2 \in G$,

$$\rho_1 \ast \rho_2 \neq \rho_2 \ast \rho_1 \neq L_{g_1} \rho_1 \ast R_{g_2} \rho_2 \neq C_{g_1} \rho_1 \ast C_{g_2} \rho_2,$$

however, entropy satisfies the following equalities

$$S(\rho_1 \ast \rho_2) = S(\rho_2 \ast \rho_1') = S(L_{g_1} \rho_1 \ast R_{g_2} \rho_2) = S(C_{g_1} \rho_1 \ast C_{g_2} \rho_2).$$

**Proof:** Each equality is proven by changing variables and using the unimodularity property in (21).
Let $F[\rho] = -\rho \log \rho$. Then due to (21), the integral over $G$ of $F[\rho(g^{-1})]$ must be the same as $F[\rho(g)]$, proving the first equality in (59). The second equality follows from the fact that $(L_{g_1} \rho_1 \ast R_{g_2} \rho_2)(g) = (\rho_1 \ast \rho_2)(g_1 \circ g \circ g_2)$ and the integral of $F[\rho(g_1 \circ g \circ g_2)]$ must be the same as $F[\rho(g)]$, again due to (21). The final equality follows in a similar way from the fact that $(C_{g_1} \rho_1 \ast C_{g_2} \rho_2)(g) = (\rho_1 \ast \rho_2)(g_1^{-1} \circ g \circ g_1)$.

Note that the equalities in (59) can be combined. For example, $S(g_1 \rho_1 \ast g_2 \rho_2) = S(L_{g_1} \rho_1 \ast R_{g_2} \rho_2) = S(C_{g_1} \rho_1 \ast C_{g_2} \rho_2)$.

**Theorem 3.8:** The equality $S(g_1 \rho_1 \ast g_2 \rho_2) = S(g_2 \rho_2 \ast g_1 \rho_1)$ holds for pdfs $\rho_1(g)$ and $\rho_2(g)$ on a unimodular Lie group $G$ in the following cases: (a) $\rho_i(g)$ for $i = 1$ or $i = 2$ is a class function; (b) $\rho_i(g)$ for $i = 1, 2$ are both symmetric functions.

**Proof:** Statement (a) follows from the fact that if either $\rho_1$ or $\rho_2$ is a class function, then convolutions commute. Statement (b) follows from the first equality in (59) and the definition of a symmetric function.

**Theorem 3.9:** Given class functions $\chi_1(g)$ and $\chi_2(g)$ that are pdfs, then for general $g_1, g_2 \in G$,

$$(\chi_1 \ast \chi_2)(g) \neq (L_{g_1} \chi_1 \ast L_{g_2} \chi_2)(g) \neq (R_{g_1} \chi_1 \ast R_{g_2} \chi_2)(g) \neq (R_{g_1} \chi_1 \ast L_{g_2} \chi_2)(g)$$

and yet

$$S(\chi_1 \ast \chi_2) = S(L_{g_1} \chi_1 \ast L_{g_2} \chi_2) = S(R_{g_1} \chi_1 \ast R_{g_2} \chi_2) = S(R_{g_1} \chi_1 \ast L_{g_2} \chi_2). \quad (60)$$

**Proof:**

Here the first and final equality will be proven. The middle one follows in the same way.

$$(L_{g_1} \chi_1 \ast L_{g_2} \chi_2)(g) = \int_G (L_{g_1} \chi_1)(h) \ast (L_{g_2} \chi_2)(g^{-1} \circ h \circ g) dh = \int_G \chi_1(g_1^{-1} \circ h \circ g) \chi_2(g_2^{-1} \circ h^{-1} \circ g) dh$$

$$= \int_G \chi_1(k) \chi_2(g_2^{-1} \circ k^{-1} \circ g_1^{-1} \circ g) dk = \int_G \chi_1(k) \chi_2(k^{-1} \circ g_1^{-1} \circ g \circ g_2^{-1}) dk$$

$$= (\chi_1 \ast \chi_2)(g^{-1} \circ g \circ g_2^{-1}).$$

Similarly,

$$(R_{g_1} \chi_1 \ast L_{g_2} \chi_2)(g) = \int_G (R_{g_1} \chi_1)(h) \ast (L_{g_2} \chi_2)(h^{-1} \circ g) dh = \int_G \chi_1(h \circ g_1) \chi_2(g_2^{-1} \circ h^{-1} \circ g) dh$$

$$= \int_G \chi_1(k) \ast \chi_2(g_2^{-1} \circ g_1 \circ k^{-1} \circ g) dk = \int_G \chi_1(k) \ast \chi_2(k^{-1} \circ g \circ g_2^{-1} \circ g_1) dk$$

$$= (\chi_1 \ast \chi_2)(g \circ g_2^{-1} \circ g_1).$$

### 4 Fisher Information and Diffusions on Lie Groups

The natural extension of the Fisher information matrix for the case when $f(g, \theta)$ is a parametric distribution on a Lie group is

$$F_{ij}(f, \theta) = \int_G \frac{1}{f} \frac{\partial f}{\partial \theta_i} \frac{\partial f}{\partial \theta_j} dg. \quad (61)$$

In the case when $\theta$ parameterizes $G$ as $g(\theta) = \exp(\sum_i \theta_i X_i)$ and $f(g, \theta) = f(g \circ \exp(\sum_i \theta_i X_i))$, then

$$\frac{\partial f}{\partial \theta_i} \bigg|_{\theta = 0} = \dot{X}_i^\ast f$$
and $F_{ij}(f,0)$ becomes
\[
F_{ij}^t(f) = \int_G \frac{1}{f} (\tilde{X}_i^t f)(\tilde{X}_j^t f)dg.
\] (62)

In a similar way, we can define
\[
F_{ij}^t(f) = \int_G \frac{1}{f} (\tilde{X}_i^t f)(\tilde{X}_j^t f)dg.
\] (63)

**Theorem 4.1:** The matrices \[ and \] have the properties
\[
F_{ij}^t(L(h)f) = F_{ij}^t(f) \quad \text{and} \quad F_{ij}^t(R(h)f) = F_{ij}^t(f)
\] (64)
and
\[
F_{ij}^t(I(f)) = F_{ij}^t(f) \quad \text{and} \quad F_{ij}^t(I(f)) = F_{ij}^t(f)
\] (65)
where $(L(h)f)(g) = f(h^{-1} \circ g)$, $(R(h)f)(g) = f(g \circ h)$, and $I(f)(g) = f(g^{-1})$.

**Proof:** The operators $\tilde{X}_i^t$ and $R(h)$ commute, and likewise $\tilde{X}_i^t$ and $L(h)$ commute. This together with the invariance of integration under shifts proves (64). From the definitions of $\tilde{X}_i^t$ and $\tilde{X}_i^t$ in (22), it follows that
\[
\tilde{X}_i^t(I(f))(g) = \left(\frac{d}{dt} f([g \circ \exp(tX_i)]^{-1})\right)_{t=0} = \left(\frac{d}{dt} f(\exp(-tX_i) \circ g^{-1})\right)_{t=0} = (\tilde{X}_i^t f)(g^{-1}).
\]

Using the invariance of integration under shifts then gives (65). As a special case, when $f(g)$ is a symmetric function, the left and right Fisher information matrices will be the same.

Note that the entries of Fisher matrices $F_{ij}^t(f)$ and $F_{ij}^t(f)$ implicitly depend on the choice of orthonormal Lie algebra basis \{X_i\}, and so it would be more descriptive to use the notation $F_{ij}^t(f,X)$ and $F_{ij}^t(f,X)$.

If a different orthonormal basis \{Y_i\} is used, such that $X_i = \sum_k a_{ik} Y_k$, then the orthonormality of both \{X_i\} and \{Y_i\} forces $A = [a_{ij}]$ to be an orthogonal matrix. Furthermore, the linearity of the Lie derivative,
\[
\tilde{X}_i f = \sum_k x_k \tilde{X}_i^t f \quad \text{where} \quad X = \sum_i x_i X_i,
\]
means that
\[
F_{ij}^t(f,X) = \int_G \frac{1}{f} \left( \sum_k a_{ik} \tilde{Y}_k^t f \right) \left( \sum_l a_{jl} \tilde{Y}_l^t f \right) dg = \sum_{k,l} a_{ik} a_{jl} F_{kl}^t(f,Y).
\]
The same holds for $F_{ij}^t$. Summarizing these results in matrix form:
\[
F^t(f,X) = A F^t(f,Y) A^T \quad \text{and} \quad F^t(f,X) = A F^t(f,Y) A^T \quad \text{where} \quad e_i^T A e_j = (X_i,Y_j).
\] (66)

This means that the eigenvalues of the Fisher information matrix (and therefore its trace) are invariant under change of orthonormal basis.

### 4.1 Fisher Information and Convolution on Groups

The decrease of Fisher information as a result of convolution can be studied in much the same way as for pdfs on Euclidean space. Two approaches are taken here. First, a straightforward application of the Cauchy-Bunyakovsky-Schwarz (CBS) inequality is used together with the bi-invariance of the integral over a unimodular Lie group to produce a bound on the Fisher information of the convolution of two probability densities. Then, a tighter bound is obtained using the concept of conditional expectation in the special case when the pdfs commute under convolution.
Theorem 4.2: The following inequalities hold for the diagonal entries of the left and right Fisher information matrices:

\[ F_{ii}^u(f_1 \ast f_2) \leq \min \{ F_{ii}^u(f_1), F_{ii}^u(f_2) \} \quad \text{and} \quad F_{ii}^l(f_1 \ast f_2) \leq \min \{ F_{ii}^l(f_1), F_{ii}^l(f_2) \}. \]  \(67\)

**Proof:** The CBS inequality holds for groups:

\[ \left( \int_G a(g)b(g)dg \right)^2 \leq \int_G a^2(g)dg \int_G b^2(g)dg. \]

If \( a(g) \geq 0 \) for all values of \( g \), then it is possible to define \( j(g) = |a(g)|^{1/2} \) and \( k(g) = |a(g)|^{1/2}b(g) \), and since \( j(g)k(g) = a(g)b(g) \),

\[ \left( \int_G a(g)b(g)dt \right)^2 \leq \left( \int_G j^2(g)dg \right) \left( \int_G k^2(t)dg \right) \]

\[ = \left( \int_G a(g)dg \right) \left( \int_G a(g)\{b(g)\}^2dg \right). \]  \(68\)

Using this version of the CBS inequality, and letting \( b(g) = \tilde{X}_{i}^l f_2(h^{-1} \circ g)/[f_2(h^{-1} \circ g)] \) and \( a(g) = f_1(h)f_2(h^{-1} \circ g) \), essentially the same manipulations as in \(10\) can be used, with the roles of \( f_1 \) and \( f_2 \) interchanged due to the fact that in general for convolution on a Lie group \((f_1 \ast f_2)(g) \neq (f_2 \ast f_1)(g)\):

\[ F_{ii}^u(f_1 \ast f_2) = \int_G \left( \int_G [\tilde{X}_{i}^l f_2(h^{-1} \circ g)]/[f_2(h^{-1} \circ g)] \cdot [f_2(h^{-1} \circ g)f_1(h)]dh \right)^2 dg \]

\[ \leq \int_G \left( \int_G [\tilde{X}_{i}^l f_2(h^{-1} \circ g)]/[f_2(h^{-1} \circ g)]^2[f_2(h^{-1} \circ g)f_1(h)]dh \right) \left( \int_G f_2(h^{-1} \circ g)f_1(h)dh \right) \]

\[ = \int_G \left( \int_G \{[\tilde{X}_{i}^l f_2(h^{-1} \circ g)]^2/f_2(h^{-1} \circ g)\}f_1(h)dh \right) \]

\[ = \int_G \left( \int_G \{[\tilde{X}_{i}^l f_2(h^{-1} \circ g)]^2/f_2(h^{-1} \circ g)\}f_1(h)dh \right) \]

\[ = F_{ii}^u(f_2) \int_G f_1(h)dh \]

\[ = F_{ii}^u(f_2) \]

Since for a unimodular Lie group it is possible to perform changes of variables and inversion of the variable of integration without affecting the value of an integral, the convolution can be written in the following equivalent ways,

\[ (f_1 \ast f_2)(g) = \int_G f_1(h)f_2(h^{-1} \circ g)dh \]  \(69\)

\[ = \int_G f_1(g \circ h^{-1})f_2(h)dh \]  \(70\)

\[ = \int_G f_1(g \circ h)f_2(h^{-1})dh \]  \(71\)

\[ = \int_G f_1(h^{-1})f_2(h \circ g)dh \]  \(72\)

It then follows that using \(70\) and the bi-invariance of integration that \(67\) holds.
4.1.1 A Tighter Bound Using Conditional Expectation for Commuting PDFs

In this subsection a better inequality is derived.

**Theorem 4.3:** The following inequality holds for the right and left Fisher information matrices:

\[
\text{tr}[F^r(\rho_1 \ast \rho_2)P] \leq \text{tr}[F^r(\rho_1)P] \quad \text{and} \quad \text{tr}[F^l(\rho_1 \ast \rho_2)P] \leq \text{tr}[F^l(\rho_1)P]
\]

where \(i = 1, 2\) and \(P\) is an arbitrary symmetric positive definite matrix with the same dimensions as \(F\).

**Proof:** Let

\[
f_{12}(h, g) = \rho_1(h)\rho_2(h^{-1} \circ g).
\]

Then

\[
f_1(h) = \int_G f_{12}(h, g) dg = \rho_1(h) \quad \text{and} \quad f_2(g) = \int_G f_{12}(h, g) dh = (\rho_1 \ast \rho_2)(g).
\]

It follows that

\[
(\tilde{X}_i f_2)(g) = \int_G \rho_1(h)\tilde{X}_i \rho_2(h^{-1} \circ g) dh.
\]

Then by the change of variables \(k = h^{-1} \circ g\),

\[
(\tilde{X}_i f_2)(g) = \int_G \rho_1(g \circ k^{-1})\tilde{X}_i \rho_2(k) dk.
\]

This means that

\[
\frac{(\tilde{X}_i f_2)(g)}{f_2(g)} = \int_G \frac{\tilde{X}_i \rho_2(k)}{\rho_2(k)} \frac{\rho_1(g \circ k^{-1})\rho_2(k)}{f_2(g)} dk = \left\langle \frac{\tilde{X}_i \rho_2(k)}{\rho_2(k)} \right\rangle g
\]

And therefore,

\[
F^r_{ii}(f_2) = \left\langle \left( \frac{\tilde{X}_i \rho_2(g)}{f_2(g)} \right)^2 \right\rangle = \left\langle \left( \frac{\tilde{X}_i \rho_2(k)}{\rho_2(k)} \right)^2 \right\rangle g \leq \left\langle \left( \frac{\tilde{X}_i \rho_2(k)}{\rho_2(k)} \right)^2 \right\rangle g = F^r_{ii}(\rho_2).
\]

An analogous argument using \(f_{12}(h, g) = \rho_1(g \circ h^{-1})\rho_2(h)\) and \(f_2(g) = (\rho_1 \ast \rho_2)(g)\) shows that

\[
\frac{(\tilde{X}_i f_2)(g)}{f_2(g)} = \left\langle \frac{\tilde{X}_i \rho_1(k)}{\rho_1(k)} \right\rangle g
\]

and

\[
F^l_{ii}(f_2) \leq F^l_{ii}(\rho_1).
\]

The above results can be written concisely by introducing an arbitrary positive definite diagonal matrix \(\Lambda\) as follows:

\[
\text{tr}[F^r(\rho_1 \ast \rho_2)\Lambda] \leq \text{tr}[F^r(\rho_1)\Lambda] \quad \text{and} \quad \text{tr}[F^l(\rho_1 \ast \rho_2)\Lambda] \leq \text{tr}[F^l(\rho_1)\Lambda].
\]

If this is true in one basis, then using \(\tilde{X}_i\), the more general statement in \((73)\) must follow in another basis where \(P = P^T > 0\). Since the initial choice of basis is arbitrary, \(\tilde{X}_i\) must hold in every basis for an arbitrary positive definite matrix \(P\). This completes the proof.

In some instances, even though the group is not commutative, the functions \(\rho_1\) and \(\rho_2\) will commute. For example, if \(\rho(g \circ h) = \rho(h \circ g)\) for all \(h, g \in G\), then \((\rho \ast \rho_1)(g) = (\rho_1 \ast \rho)(g)\) for any reasonable choice of \(\rho_1(g)\). Or if \(\rho_2 = \rho_1 \ast \rho_1 \ast \cdots \rho_1\) it will clearly be the case that
$\rho_1 \ast \rho_2 = \rho_2 \ast \rho_1$. If, for whatever reason, $\rho_1 \ast \rho_2 = \rho_2 \ast \rho_1$ then (73) can be rewritten in the following form:

$$
\text{tr}[F^r(\rho_1 \ast \rho_2)P] \leq \min\{\text{tr}[F^r(\rho_1)P], \text{tr}[F^r(\rho_2)P]\}
$$

and

$$
\text{tr}[F^l(\rho_1 \ast \rho_2)P] \leq \min\{\text{tr}[F^l(\rho_1)P], \text{tr}[F^l(\rho_2)P]\}
$$

(76)

Theorem 4.4: When $\rho_1 \ast \rho_2 = \rho_2 \ast \rho_1$ the following equality holds

$$
\frac{1}{\text{tr}[F^r(\rho_1 \ast \rho_2)P]} \leq \frac{1}{\text{tr}[F^r(\rho_1)P]} + \frac{1}{\text{tr}[F^r(\rho_2)P]} \quad \text{for any} \quad P = P^T > 0,
$$

(77)

and likewise for $F^l$.

Proof: Returning to (74) and (75), in the case when $\rho_1 \ast \rho_2 = \rho_2 \ast \rho_1$ it is possible to write

$$
\langle \frac{\tilde{X}_i^r f_2(g)}{f_2(g)} \rangle = \left\langle \frac{\langle \tilde{X}_i^r \rho_1(k) \rangle_{\rho_2(k)}}{\rho_1(k)} \right\rangle = \left\langle \frac{\langle \tilde{X}_i^r \rho_1(k) \rangle_{\rho_1(k)}}{\rho_1(k)} \right\rangle
$$

(78)

and

$$
\langle \frac{\tilde{X}_i^l f_2(g)}{f_2(g)} \rangle = \left\langle \frac{\langle \tilde{X}_i^l \rho_1(k) \rangle_{\rho_1(k)}}{\rho_1(k)} \right\rangle = \left\langle \frac{\langle \tilde{X}_i^l \rho_1(k) \rangle_{\rho_1(k)}}{\rho_1(k)} \right\rangle
$$

Since the following calculation works the same way for both the ‘l’ and ‘r’ cases, consider only the ‘r’ case for now. Multiplying the first equality in (78) by $1 - \beta$ and the second by $\beta$ and adding together$^1$

$$
\langle \frac{\tilde{X}_i^r f_2(g)}{f_2(g)} \rangle = \beta \left\langle \frac{\langle \tilde{X}_i^r \rho_1(k) \rangle_{\rho_1(k)}}{\rho_1(k)} \right\rangle + (1 - \beta) \left\langle \frac{\langle \tilde{X}_i^r \rho_2(k') \rangle_{\rho_2(k')}}{\rho_2(k')} \right\rangle
$$

for arbitrary $\beta \in [0, 1]$.

Now squaring both sides and taking the (unconditional) expectation, and using Jensen’s inequality yields:

$$
\left\langle \left\langle \frac{\tilde{X}_i^r f_2(g)}{f_2(g)} \right\rangle^2 \right\rangle = \beta^2 \left\langle \left\langle \frac{\langle \tilde{X}_i^r \rho_1(k) \rangle_{\rho_1(k)}}{\rho_1(k)} \right\rangle^2 \right\rangle + (1 - \beta)^2 \left\langle \left\langle \frac{\langle \tilde{X}_i^r \rho_2(k') \rangle_{\rho_2(k')}}{\rho_2(k')} \right\rangle^2 \right\rangle
$$

(79)

This statement simply says

$$
F_{\rho_1}(\rho_1 \ast \rho_2) \leq \beta^2 F_{\rho_1}(\rho_1) + (1 - \beta)^2 F_{\rho_2}(\rho_2).
$$

(80)

The value of $\beta \in [0, 1]$ that gives the tightest bound is

$$
\beta = \frac{F_{\rho_2}(\rho_2)}{F_{\rho_1}(\rho_1) + F_{\rho_2}(\rho_2)}.
$$

$^1$The names of the dummy variables $k$ and $k'$ are unimportant. However, at this stage it is important that the names be different in order to emphasize their statistical independence.
resulting in the inequality
\[
\frac{1}{F_{ii}^r(\rho_1 \ast \rho_2)} \leq \frac{1}{F_{ii}^r(\rho_1)} + \frac{1}{F_{ii}^r(\rho_2)}.
\] (81)

Alternatively, if before computing the optimal \(\beta\) we first multiply both sides of (80) by \(\lambda_i\) and sum over \(i\), the result will be
\[
\text{tr}[F^r(\rho_1 \ast \rho_2)\Lambda] \leq \beta \text{tr}[F^r(\rho_1)\Lambda] + (1 - \beta)^2 \text{tr}[F^r(\rho_2)\Lambda].
\]

Again, since the basis is arbitrary, \(\Lambda\) can be replaced with \(P\). Then the optimal value of \(\beta\) will give (77).

4.1.2 A Special Case: \(SO(3)\)

Consider the group of \(3 \times 3\) orthogonal matrices with determinant +1. Let \(\tilde{X}^r = [\tilde{X}_1^r, \tilde{X}_2^r, \tilde{X}_3^r]^T\) and \(\tilde{X}^l = [\tilde{X}_1^l, \tilde{X}_2^l, \tilde{X}_3^l]^T\). These two gradient vectors are related to each other by an adjoint matrix, which for this group is a rotation matrix [24]. Therefore, in the case when \(G = SO(3)\),
\[
\|\tilde{X}^r f\|_2 = \|\tilde{X}^l f\|_2 \implies \text{tr}[F^r(f)] = \text{tr}[F^l(f)]
\]

Therefore, the inequalities in (80) will hold for pdfs on \(SO(3)\) regardless of whether or not the functions commute under convolution, but restricted to the condition \(P = I\).

5 Generalizing the de Bruijn Identity to Lie Groups

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

**Theorem 5.1:** Let \(f_{D,h,t}(g) = f(g,t;D,h)\) denote the solution to the diffusion equation (29) with constant \(h\) subject to the initial condition \(f(g,0;D,h) = \delta(g)\). Then for any well-behaved pdf \(\alpha(g)\),
\[
\frac{d}{dt}S(\alpha \ast f_{D,h,t}) = \frac{1}{2} \text{tr}[DF^r(\alpha \ast f_{D,h,t})].
\] (82)

**Proof:** It is easy to see that the solution of the diffusion equation
\[
\frac{\partial \rho}{\partial t} = \frac{1}{2} \sum_{i,j=1}^n D_{ij} \tilde{X}^r_i \tilde{X}^r_j \rho - \sum_{k=1}^n h_k \tilde{X}^r_k \rho
\] (83)

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

Taking the time derivative of \(S(\rho(g,t))\) we get
\[
\frac{d}{dt}S(\rho) = -\frac{d}{dt} \int_G \rho(g,t) \log \rho(g,t) dg = - \int_G \left[ \frac{\partial \rho}{\partial t} \log \rho + \frac{\partial \rho}{\partial t} \right] dg.
\] (84)

Using (83), the partial with respect to time can be replaced with Lie derivatives. But
\[
\int_G \tilde{X}^r_k \rho dg = \int_G \tilde{X}^r_i \tilde{X}^r_j \rho dg = 0,
\]
so the second term on the right side of (84) completely disappears. Using the integration-by-parts formula\footnote{There are no surface terms because, like the circle and real line, each coordinate in the integral either wraps around or goes to infinity.}
\[
\int_G f_1 \tilde{X}^r_k f_2 dg = - \int_G f_2 \tilde{X}^r_k f_1 dg,
\]
with \( f_1 = \log \rho \) and \( f_2 = \rho \) then gives

\[
\frac{d}{dt} S(\alpha * f_{D,h,t}) = \frac{1}{2} \sum_{i,j=1}^{n} D_{ij} \int_{G} \frac{1}{\alpha * f_{D,h,t}} \tilde{X}_{i}^{\tau}(\alpha * f_{D,h,t}) \tilde{X}_{j}^{\tau}(\alpha * f_{D,h,t}) \, dg \\
= \frac{1}{2} \sum_{i,j=1}^{n} D_{ij} F_{ij}(\alpha * f_{D,h,t}) \\
= \frac{1}{2} \text{tr}[DF^\tau(\alpha * f_{D,h,t})].
\]

The implication of this is that

\[
S(\alpha * f_{D,h,t_2}) - S(\alpha * f_{D,h,t_1}) = \frac{1}{2} \int_{t_1}^{t_2} \text{tr}[DF^\tau(\alpha * f_{D,h,t})] \, dt
\]

6 Information-Theoretic Inequalities from Log-Sobolev Inequalities

In this section information-theoretic identities are derived from Log-Sobolev inequalities. Subsection 5.1 provides a brief review of Log-Sobolev inequalities. Subsection 6.2 then uses these to write information-theoretic inequalities.

6.1 Log-Sobolev Inequalities in \( \mathbb{R}^n \) and on Lie Groups

The log-Sobolev inequality can be stated as \([7, 8, 50]\):

\[
\int_{\mathbb{R}^n} |\psi(x)|^2 \log |\psi(x)|^2 \, dx \leq \frac{n}{2} \log \left[ \frac{2}{\pi e n} \int_{\mathbb{R}^n} |\nabla \psi|^2 \, dx \right]
\] (85)

where

\[
\nabla \psi = \left[ \frac{\partial \psi}{\partial x_1}, \ldots, \frac{\partial \psi}{\partial x_n} \right]^T \quad \text{and} \quad \int_{\mathbb{R}^n} |\psi(x)|^2 \, dx = 1.
\]

Here \( \log = \log_{e} \). Actually, there is a whole family of log-Sobolev inequalities, and \([50]\) represents the tightest of these. The original form of the log-Sobolev inequality as introduced by Gross in \([33]\) is

\[
\frac{1}{2} \int_{\mathbb{R}^n} |\phi(x)|^2 \log |\phi(x)|^2 \rho(x) \, dx \leq \int_{\mathbb{R}^n} |\nabla \phi(x)|^2 \rho(x) \, dx + \|\phi\|_{L^2(\mathbb{R}^n, \rho)}^2 \log \|\phi\|_{L^2(\mathbb{R}^n, \rho)}^2
\] (86)

where

\[
\|\phi\|_{L^2(\mathbb{R}^n, \rho)}^2 = \int_{\mathbb{R}^n} |\phi(x)|^2 \rho(x) \, dx.
\]

Here \( \rho(x) = \rho(x, 0) = (2\pi)^{-n/2} \exp(-\|x\|^2/2) \) is the solution to the heat equation on \( \mathbb{R}^n \) evaluated at \( t = 1 \).

Several different variations exist. For example, by rescaling, it is possible to rewrite \([50]\) with \( \rho(x, t) \) in place of \( \rho(x) \) by introducing a multiplicative factor of \( t \) in the first term on the right hand side of the equation. Or, by letting \( \phi(x) = \rho^{-\frac{a}{2}}(x) \psi(x/\alpha) \) for some scaling factor \( a > 0 \), substituting into \([50]\), and integrating by parts then gives \([50]\)

\[
\int_{\mathbb{R}^n} |\psi(x)|^2 \log \frac{|\psi(x)|^2}{\|\psi\|_2^2} \, dx + n(1 + \log \alpha) \|\psi\|_2^2 \leq \frac{a^2}{\alpha} \int_{\mathbb{R}^n} |\nabla \psi(x)|^2 \, dx
\]

where

\[
\|\psi\|_2^2 = \int_{\mathbb{R}^n} |\psi(x)|^2 \, dx \quad \text{and} \quad \|\nabla \psi(x)|^2 = \nabla \psi(x) \cdot \nabla \psi(x).
\]
This, together with an optimization over \( a \) gives \( \text{(85)} \).

Gross subsequently extended \( \text{(85)} \) to Lie groups \( \text{(87)} \) as

\[
\int_G \left\{ |\phi(g)|^2 \log |\phi(g)| \right\} \rho(g, t) \, dg \leq C_G(t) \int_G \|\tilde{X}^T \phi(g)\|^2 \rho(g, t) \, dg + \|\phi\|^2_{\text{L}^2(G, \rho_t)} \log \|\phi\|^2_{\text{L}^2(G, \rho_t)}
\]

where \( \rho(g, t) \) is the solution to the diffusion equation in \( \text{(85)} \) with \( h_i = 0, D_{ij} = \delta_{ij} \), initial condition \( \rho(g, 0) = \delta(g) \), and

\[
\tilde{X}^T \phi = [\tilde{X}_1^T \phi, ..., \tilde{X}_n^T \phi]^T \quad \text{and} \quad \|\phi\|^2_{\text{L}^2(G, \rho_t)} = \int_G |\phi(g)|^2 \rho(g, t) \, dg.
\]

In \( \text{(88)} \) the scalar function \( C_G(t) \) depends on the particular group. For \( G = (\mathbb{R}^n, +) \) we have \( c_{\mathbb{R}^n}(t) = t \), and likewise \( c_{SO(n)}(t) = t \).

In analogy with the way that \( \text{(88)} \) evolved from \( \text{(87)} \), a descendant of \( \text{(88)} \) for noncompact unimodular Lie groups is \( \text{(89)} \).

\[
\int_G |\psi(g)|^2 \log |\psi(g)|^2 \, dg \leq \frac{n}{2} \log \left[ C_G \frac{2}{\pi e n} \int_G \|\tilde{X} \psi\|^2 \, dg \right]
\]

The only difference is that, to the author’s knowledge, the sharp factor \( C_G \) in this expression is not known for most Lie groups. The information-theoretic interpretation of these inequalities is provided in the following subsection.

### 6.2 Information-Theoretic Inequalities

For our purposes the form in \( \text{(89)} \) will be most useful. It is interesting to note in passing that Beckner has extended this inequality to the case where the domain, rather than being \( \mathbb{R}^n \), is the hyperbolic space \( \mathbb{H}^n \cong SL(2, \mathbb{R})/SO(2) \) and the Heisenberg groups \( H(n) \), including \( H(1) \), \( \mathbb{H} \). Our goal here is to provide an information-theoretic interpretation of the inequalities from the previous section.

**Theorem 6.1:** Entropy powers and Fisher information are related as

\[
[N(f)]^{-1} \leq \frac{1}{n} \text{tr}(F) \quad \text{where} \quad N(f) = C_G \frac{2}{\pi e n} \exp \left[ \frac{2}{n} \text{tr}(F) \right].
\]

**Proof:** We begin by proving \( \text{(89)} \) for \( G = (\mathbb{R}^n, +) \). Making the simple substitution \( f(x) = |\psi(x)|^2 \) into \( \text{(89)} \) and requiring that \( f(x) \) be a pdf gives

\[
\int_{\mathbb{R}^n} f(x) \log f(x) \, dx \leq \frac{n}{2} \log \left[ \frac{1}{2 \pi e n} \int_{\mathbb{R}^n} \frac{1}{\|\nabla f\|^2} \, dx \right].
\]

or

\[
- S(f) \leq \frac{n}{2} \log \frac{\text{tr}(F)}{2 \pi e n} \implies \exp \left[ - \frac{2}{n} S(f) \right] \leq \frac{\text{tr}(F)}{2 \pi e n} \implies [N(f)]^{-1} \leq \frac{1}{n} \text{tr}(F).
\]

Here \( S(f) \) is the Boltzmann-Shannon entropy of \( f \) and \( F \) is the Fisher information matrix. As is customary in information theory, the entropy power can be defined as \( N(f) \) in \( \text{(89)} \) with \( C_G = 1 \). Then the log-Sobolev inequality in the form in \( \text{(89)} \) is written as \( \text{(89)} \).

For the more general case, starting with \( \text{(89)} \) and letting \( f(g) = |\tilde{\psi}(g)|^2 \) gives

\[
\int_G f(g) \log f(g) \, dg \leq \frac{n}{2} \log \left[ C_G \frac{2}{\pi e n} \int_G \frac{1}{\|\tilde{X} \tilde{f}\|^2} \, dg \right] \implies - S(f) \leq \frac{n}{2} \log \left[ C_G \frac{2}{\pi e n} \text{tr}(F) \right].
\]

The rest is the same as for the case of \( \mathbb{R}^n \).

Starting with Gross’s original form of log-Sobolev inequalities involving the heat kernel, the following information-theoretic inequality results:
Theorem 6.2: The Kullback-Leibler divergence and Fisher-Information distance of any arbitrary pdf and the heat kernel are related as

\[ D_{KL}(f \parallel \rho_t) \leq \frac{c_G(t)}{2} D_{F I}(f \parallel \rho_t) \]  

where in general given \( f_1(g) \) and \( f_2(g) \),

\[ D_{F I}(f_1 \parallel f_2) = \int_G \left\| \frac{1}{f_1} \tilde{X} f_1 - \frac{1}{f_2} \tilde{X} f_2 \right\|^2 f_1 dg. \]

Proof: Starting with (87), let \( \psi(g,t) = [\rho(g,t)]^{-\frac{1}{2}} f(g) \frac{1}{2} \) where \( f(g) \) is a pdf. Then

\[ \int_G |\psi(g,t)|^2 \rho(g,t) dg = \int_G f(g) dg = 1 \]

and so \( \log \| \phi \|_L^2(G,\rho_1) = 0 \), and we have

\[ \frac{1}{2} \int_G f(g) \log \frac{f(g)}{\rho(g,t)} dg \leq \int_G \| \tilde{X}(\rho(g,t))^{-\frac{1}{2}} [f(g)]^{\frac{1}{2}} \| \rho(g,t) dg. \]

By using the chain rule and product rule for differentiation,

\[ \tilde{X}(\rho(g,t))^{-\frac{1}{2}} [f(g)]^{\frac{1}{2}} = \frac{1}{2} f^{\frac{1}{2}} \tilde{X} f - \frac{1}{2} f^{\frac{1}{2}} \rho_t^{-1} \tilde{X} \rho_t. \]

Substituting into the right hand side of (87) then gives (92).

In the functional analysis community from which log-Sobolev inequalities emerged it is rarely, if ever, stated in these terms. One exception is the work of Carlen [17], which addresses Theorem 6.1 for the case of \( G = \mathbb{R}^n \). Moreover, the author has not found analogs of (90) in the context of Lie groups in the literature.

7 The Entropy-Power Inequality (or Lack Thereof)

One of the fundamental inequalities of information theory is the entropy power inequality

\[ N(f_1 * f_2) \geq N(f_1) + N(f_2) \]

for any pdfs \( f_1 \) and \( f_2 \) on \( \mathbb{R}^n \) with \( N(f_1) \) defined as in (80) for \( C_{\mathbb{R}^n} = 1 \). This was first stated by Shannon [63] together with a verification of the necessary conditions for it to be true. This was followed up with proofs of sufficiency by Stam and Blachman [12, 69]. Without going into too many details, the key technical points of their proofs require two properties. First,

\[ f_1 * \rho_{t_1} * f_2 * \rho_{t_2} = f_1 * f_2 * \rho_{t_1} * \rho_{t_2} \]

(which is not a problem in \( \mathbb{R}^n \) since convolution is commutative). Second, they also use a scaling argument requiring that any pdf \( f(x) \) that is scaled as \( f_s(x) = s \cdot f(s \cdot x) \) will become the Dirac delta function as \( s \to 0 \). That is not to say that these two properties are essential to proving the entropy power inequality, but rather only that they are the properties that are used in the most familiar proofs.

However, there is somewhat of a conundrum because for compact Lie groups, the heat kernel \( \rho_t(g) \) is a class function, and therefore satisfies the first condition. However, there is no natural way to rescale on a compact Lie group (not even on the circle group, \( SO(2) \)). And in fact, it is easy to see that on compact Lie groups the entropy power inequality does not hold. For example, the limiting distribution on a compact Lie group is \( \rho_\infty = 1 \) with entropy \( S(\rho_\infty) = 0 \), and entropy power \( N(\rho_\infty) = 1 \). Since \( \rho_\infty * f = \rho_\infty \) for any pdf, \( f \), we get \( N(\rho_\infty * f) = 1 < 1 + N(f) \) since \( N(f) > 0 \) always.
On the other hand, it is possible for some groups to introduce a concept of scaling. For example, it is possible to do this in the Heisenberg group, roughly speaking, because all coordinate directions extend to infinity. Groups that admit a scaling property have been studied extensively [31]. However, whether the heat equations on such groups yield solutions that are class functions then becomes an issue. Regardless, for the groups of primary interest in engineering applications, i.e., the rotation and rigid-body motion groups, the possibilities for an entropy power inequality appear to be pretty slim.

8 Conclusions

By collecting and reinterpreting results relating to the study of diffusion processes, harmonic analysis, and log-Sobolev inequalities on Lie groups, and merging these results with new definitions of covariance and Fisher information, many inequalities of information theory were extended here to the context of probability densities on unimodular Lie groups. In addition, the natural decomposition of groups into cosets, double cosets, and the nesting of subgroups provides some inequalities that result from the Kullback-Leibler divergence of probability densities on Lie groups. Some special inequalities related to finite groups were also provided.

While the emphasis of this paper was on the discovery of fundamental inequalities and the introduction of Lie group concepts to the information theory audience, the motivation for this study originated with applications in robotics and other areas. Though these applications were not explored here, references to the literature pertaining to robot motion and image reconstruction were provided.

Acknowledgments

This work was performed under support from the NIH Grant R01 GM075310 “Group Theoretic Methods in Protein Structure Determination.”

References

[1] Adler, R. L., Konheim, A. G., McAndrew, M. H., “Topological Entropy,” Transactions of the American Mathematical Society, 114(2):309-319, 1965.
[2] Amari, S., Nagaoka, H., Methods of Information Geometry, Translations of Mathematical Monographs 191, American Mathematical Society, Providence, RI, 2000.
[3] Arnol’d, V.I., Avez, A., Ergodic Problems of Classical Mechanics, W.A. Benjamin, New York, 1968.
[4] D. Bakry, D. Concordet, M. Ledoux, “Optimal Heat Kernel Bounds Under Logarithmic Sobolev Inequalities,” ESAIM: Probability and Statistics, Vol. 1, pp. 391-407, December 1997.
[5] Baldwin, G., Mahony, R., Trumpf, J., “A Nonlinear Observer for 6 DOF Pose Estimation from Inertial and Bearing Measurements,” IEEE International Conference on Robotics and Automation, Kobe, Japan, May, 2009.
[6] Barron, A.R., “Entropy and the central limit theorem,” Ann. Prob., 14, 336-342, 1986.
[7] W. Beckner, Sharp inequalities and geometric manifolds, J. Fourier Anal. Appl. 3 (1997), 825-836.
[8] W. Beckner, Geometric inequalities in Fourier analysis, Essays on Fourier Analysis in Honor of Elias M. Stein, Princeton University Press, 1995, pp. 36-68
[9] Berg, B. C. E. coli in Motion, Springer, New York, 2003.
[10] Billingsley, P., Ergodic Theory and Information, Robert E. Krieger Publishing Co., Huntington, New York, 1978.
[11] Birkhoff, G.D., “Proof of the ergodic theorem,” Proc Natl Acad Sci USA 17: 656660, 1931.
[12] Blachman, N.M., “The convolution inequality for entropy powers,” IEEE Trans. Inform. Theory, 11(2): 267-271, 1965.

[13] Boothroyd G., Assembly Automation and Product Design, Second Edition, CRC Press, Boca Raton, FL, 2005

[14] Brockett, R.W., “System Theory on Group Manifolds and Coset Spaces,” SIAM J. Control, Vol. 10, No. 2, pp. 265-284, May 1972.

[15] Brockett, R.W., “Lie algebras and Lie groups in control theory,” in Geometric Methods in System Theory, (D.Q. Mayne and R.W. Brockett, eds.), Reidel Publishing Company, Dordrecht-Holland, 1973.

[16] Brown, L.D., “A proof of the Central Limit Theorem motivated by the Cramér-Rao inequality,” in G. Kallianpur, P.R. Krishnaiah, and J.K. Ghosh, eds., Statistics and Probability: Essays in Honour of C.R. Rao, pp. 141-148, North-Holland, New York, 1982.

[17] Carlen, E.A., “Superadditivity of Fisher’s Information and Logarithmic Sobolev Inequalities,” Journal Of Functional Analysis, 101, 194-211 (1991)

[18] Censi, A., “On Achievable Accuracy for Pose Tracking,” IEEE International Conference on Robotics and Automation, Kobe, Japan, May, 2009.

[19] Chan, T.H., R.W., Yeung, “On a Relation Between Information Inequalities and Group Theory,” IEEE Transactions On Information Theory, 48(7):1992-1995 JULY 2002

[20] Chan, T.H., “Group characterizable entropy functions,” ISIT2007, Nice, France, June 24 - June 29, 2007, pp. 506-510.

[21] Chirikjian, G.S. “Fredholm Integral Equations on the Euclidean Motion Group.” Inverse Problems 12(Oct. 1996) 579-599.

[22] Chirikjian, G.S., Wang, Y.F., “Conformational Statistics of Stiff Macromolecules as Solutions to PDEs on the Rotation and Motion Groups,” Physical Review E, Vol. 62, No. 1, July 2000., pp. 880-892.

[23] 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.

[24] G.S. Chirikjian, A.B. Kyatkin, Engineering Applications of Noncommutative Harmonic Analysis CRC Press, Boca Raton, FL, 2001.

[25] Chirikjian, G.S., “Parts Entropy and the Principal Kinematic Formula,” IEEE Conference on Automation Science and Engineering, Washington D.C., August 2008.

[26] Cover, T.M., Thomas, J.A., Elements of Information Theory, Wiley-Interscience, 2nd ed., Hoboken, NJ, 2006.

[27] Dembo, A., Cover, T.M., Thomas J.A., “Information Theoretic Inequalities,” IEEE Transactions On Information Theory 37(6):1501-1518 NOV 1991

[28] Diaconis, P., Group Representations in Probability and Statistics, Lecture Notes-Monograph Series, S.S. Gupta Series Editor, Institute of Mathematical Statistics, Hayward, CA 1988.

[29] T. E. Duncan, “An Estimation problem in compact Lie groups,” Syst. Control Lett. 10, 257-263 (1998).

[30] Fisher, R.A., “Theory of statistical estimation” Proc. Cambridge Phil. Soc. 22, 700-725, 1925.

[31] Folland, G.B., Stein, E.M., Hardy Spaces on Homogeneous Groups, Mathematical Notes 28, Princeton University Press, Princeton, NJ, 1982.

[32] I. M. Gelfand, R. A. Minlos, Z. Ya. Shapiro: Representations of the Rotation and Lorentz Groups and Their Applications, Pergamon Press, New York, 1963.
[33] L. Gross, “Logarithmic Sobolev inequalities,” *Amer. J. Math.* 97 (1975), 1061-1083.
[34] L. Gross, “Logarithmic Sobolev inequalities on Lie groups,” *Illinois J. Math.*, 36(3), Fall 1992, pp. 447-490.
[35] Grenander, U., *Probabilities on Algebraic Structures*, Dover Edition, 2008. (originally, Wiley, 1963).
[36] D. Gurari: *Symmetry and Laplacians. Introduction to Harmonic Analysis, Group Representations and Applications*, Elsevier Science Publisher, The Netherlands, 1992.
[37] Helgason, S., *Groups and Geometric Analysis*, American Mathematical Society, Providence, RI, 2000.
[38] Heyer, H., *Probability Measures on Locally Compact Groups*, Springer-Verlag, New York, 1977.
[39] Halmos, P.R., *Lectures on Ergodic Theory*, The Mathematical Society of Japan, Tokyo, 1956.
[40] Johnson, O., Suhov, Y., “Entropy and Convergence on Compact Groups,” *Journal of Theoretical Probability*, Vol. 13, No. 3, 2000, pp. 843-857.
[41] Johnson, O., *Information Theory and the Central Limit Theorem*, Imperial College Press, London, 2004.
[42] Jurdjevic, V., Sussmann, H.J., “Control Systems on Lie Groups,” *Journal of Differential Equations*, Vol. 12, pp. 313-329, 1972.
[43] Khinchin, A.I., “Zur Birkhoff’s Lösung des Ergodenproblems,” *Math. Ann.* 107: 485-488, 1932.
[44] P. T. Kim, “Deconvolution density estimation on SO(N),” *Ann. Stat.* 26(3), 10831102 (1998).
[45] Kolmogorov, A.N., “On Dynamical Systems with an Integral Invariant on the Torus,” *Dokl. Acad. Nauk. SSSR* 93: 763-766, 1953.
[46] Koo, J.-Y., Kim, P.T., “Asymptotic Minimax Bounds for Stochastic Deconvolution Over Groups,” *IEEE Transactions on Information Theory*, Volume 54(1):289 - 298, Jan. 2008.
[47] Kutzer, M.D.M., Armand, M., Lin, E., Scheidt, D., Chirikjian, G.S., “Toward Cooperative Team-Diagnosis in Multi-robot Systems,” *International Journal of Robotics Research* 27: 1069-1090, Sept. 2008.
[48] K. Lee, G.S. Chirikjian, (2007) Robotic Self-Replication from Low-Complexity Parts. *IEEE Robotics and Automation Magazine*, 14. 4 pp. 34-43 (2007)
[49] Li, H., Chong, E. K.P., “On Connections between Group Homomorphisms and the Ingleton Inequality,” *ISIT2007*, Nice, France, June 24 - June 29, 2007, pp. 1996-1999.
[50] Lieb, E.H., Loss, M., *Analysis, 2nd ed.*, American Mathematical Society, Providence, RI, 2001.
[51] Linnik, Y.V., “An information-theoretic proof of the Central Limit Theorem with the Lindeberg Condition,” *Theory of Probability and its Applications* 4(3), 288299, 1959.
[52] G. W. Mackey: *Induced Representations of Groups and Quantum Mechanics*, W. A. Benjamin, Inc., New York and Amsterdam, 1968.
[53] Major, P., Shlosman, S.B., “A local limit theorem for the convolution of probability measures on a compact connected group,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 50, 137-148, 1979.
[54] A. Makadia and K. Daniilidis, “Rotation estimation from spherical images,” *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 11701175 (2006).
[55] Malis, E., Hamel, T., Mahony, R., Morin, P., “Dynamic Estimation of Homography Transformations on the Special Linear Group for Visual Servo Control,” *IEEE International Conference on Robotics and Automation*, Kobe, Japan, May, 2009
[56] Manyika, J., Durrant-Whyte, H., “Data Fusion and Sensor Management: A Decentralized Information-Theoretic Approach,” Ellis Horwood, New York, 1994.

[57] W. Miller: *Lie Theory and Special Functions*, Academic Press, New York, 1968;

[58] Park, W., Liu, Y., Moses, M., Chirikjian, G.S., “Kinematic State Estimation and Motion Planning for Stochastic Nonholonomic Systems Using the Exponential Map,” *Robotica*, 26(4), 419-434. July-August 2008

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

[60] A.C. Sanderson, “Part Entropy Method for Robotic Assembly Design,” Proceedings of International Conference on Robotics, 1984.

[61] Shlosman, S.B., “Limit theorems of probability theory on compact topological groups,” *Theory of Probability and Its Applications*, 25, 604-609, 1980.

[62] Shlosman, S.B., “The influence of noncommutativity on limit theorems,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 65, 627-636, 1984.

[63] Shannon, C.E., Weaver, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.

[64] Sinai, Ya. G., “On the Notion of Entropy of Dynamical Systems,” *Dokl. Acad. Sci. USSR* 124(4): 768-771, 1959.

[65] P. Smith, T. Drummond and K. Roussopoulos, “Computing MAP Trajectories by Representing, Propagating and Combining PDFs over Groups,” Proceedings of the 9th IEEE International Conference on Computer Vision, vol. 2, Nice, France (2003) pp. 12751282.

[66] Stam, A.J., “Some inequalities satisfied by the quantities of information of Fisher and Shannon,” *Information and Control*, 2(2): 101-112, June 1959.

[67] Steele, J.M., “Fisher Information and Detection of a Euclidean Perturbation of an Independent Stationary Process,” *The Annals of Probability*, 14(1): 326-335, 1986.

[68] Stromberg, K., “Probabilities on a compact group,” *Trans. Amer. Math. Soc.* 94, 295-309, 1960.

[69] S. Su and C. S. G. Lee, Manipulation and propagation of uncertainty and verification of applicability of actions in assembly tasks, *IEEE Trans. Syst. Man Cybern.* 22(6), 13761389 (1992).

[70] M. Sugiura: *Unitary Representations and Harmonic Analysis*, 2nd edition, Elsevier Science Publisher, The Netherlands, 1990.

[71] M. E. Taylor, *Noncommutative Harmonic Analysis* (Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1986).

[72] Thrun, S., Burgard, W., Fox, D., *Probabilistic Robotics*, MIT Press, Cambridge, MA, 2005.

[73] Varadarajan, V.S., *An Introduction to Harmonic Analysis on Semisimple Lie Groups*, Cambridge University Press, Cambridge, England, 1989.

[74] Vergassola1, M., Villermaux, E., Shraiman, B.I., “Infotaxis as a strategy for searching without gradients,” *Nature*, Vol 445(25):406-409, January 2007.

[75] N. J. Vilenskin, A. U. Klimyk: *Representation of Lie Group and Special Functions*, Vol. 1-3, Kluwer Academic Publishers, The Netherlands, 1991.

[76] 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. 12581273

[77] Wiener, N., “The Ergodic Theorem,” *Duke Math. J.* 5: 1-18, 1939.
[78] A. S. Willsky, “Some Estimation Problems on Lie Groups,” in Geometric Methods in System Theory (D. Q. Mayne and R. W. Brockett, eds.) (Reidel Publishing Company, Dordrecht- Holland, 1973) pp. 305313.

[79] B. Yazici, Stochastic deconvolution over groups, IEEE Transactions in Information Theory, 50 (2004), 494510.

[80] Zelobenko D. P., Compact Lie Groups and their Representations, American Mathematical Society, Rhode Island, 1973.

[81] Zhang, Z., Yeung, R.W., “On characterization of entropy function via information inequalities,” IEEE Transactions On Information Theory, 44(4): 1440-1452, 1998.

[82] Zhou, Y., Chirikjian, G.S., “Conformational statistics of bent semiflexible polymers,” Journal of Chemical Physics 119 (9): 4962-4970 SEP 1 2003

[83] Zimmer, R.J., Morris, D.W., Ergodic theory, groups, and geometry American Mathematical Society, Providence, R.I., 2008.