2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

Information Theory on Lie Groups and Mobile Robotics Applications

Gregory S. Chirikjian

Abstract— **Information Theory is concerned with the reliable transmission of information through noisy environments. This relates to communicating agents, as well as one agent communicating with the environment by taking measurements (e.g., robotic sensing). Typically information theory is formulated in the context of probability on either discrete spaces or continuous Euclidean spaces in which the operation of addition makes sense. Some have extended information theory techniques to differential geometric settings. However, only in the context of group theory can the concept of addition be replaced in a meaningful way with a group operation. This paper presents concepts of information theory on Lie groups developed by the author, and illustrates their application to mobile robotics problems. In particular, the concepts of Shannon entropy, Kullback-Leibler divergence, the Cramér-Rao bound for pose data are developed, and some theorems about their properties are proved. It is also illustrated how these concepts might be integrated into pose estimation, localization, and odor-plume source detection.**

I. INTRODUCTION

Over the past decade, problems in mobile robotics have received considerable attention. Two classes of problems that both fall under the category of estimation are simultaneous localization and mapping (SLAM) [21], [9] and odor source detection [20], [19]. Both problems are probabilistic. In the former, sensors such as sonar or optical range finders are used to map the position and orientation (or "pose") of the robot in the plane or in space [14]. In the latter, the goal is to find the position of the source of an odor plume using artificial olfaction/chemical sensors.

In this paper it is shown how these two problems can be put into the same analytical framework in which information gathering is combined with stochastic models of robot sensing and locomotion capabilities. Since the space of all poses (homogeneous transformation matrices) in n-dimensional space together with the usual operation of composition (matrix multiplication) forms a Lie group (the special Euclidean group, $SE(n)$), and since the state of a slow-moving mobile robot can be described by its pose, the theory presented here has a natural Lie-group-theoretic flavor. This has been exploited in related work including [17], [11], and of course there is a huge literature on geometric mechanics, much of which is summarized in [3], as well as differential-geometric methods in computer vision [12], [13]. But to the knowledge of the author, the topic of information-theoretic methods on Lie groups has not been explored previously.

This work was supported by the National Science Foundation under grant IIS-0915542.

G. Chirikjian is with the Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA gregc@jhu.edu

For concreteness we will restrict the discussion to the case of rigid planar motions here, but the methods apply equally well to both spatial motions as well as articulated bodies, the configurations of which can be described by more complicated Lie groups. Unlike prior efforts in SLAM in which the pose of a mobile robot is treated as a vector $[x, y, \theta] \in \mathbb{R}^3$, here poses are treated as group elements $g(x, y, \theta) \in SE(2)$. Probability densities describing the pose of a robot given a set of measurements are therefore probability densities on the group $SE(2)$.

The remainder of this introduction reviews recent work on the use of information-theoretic methods in mobile robot localization and odor-plume source finding. In both problems the classical Cramér-Rao bound has been used in previous works without consideration of the properties of $SE(2)$. This can be done when robots are considered to be points that can move without consideration of nonholonomic constraints in relatively uncluttered spaces. However, if one wants to incorporate nonholonomic constraints and uncertainty in robot actuation capabilities (in addition to the uncertainty in sensing that has been addressed in prior work), then it is difficult to avoid the properties of $SE(2)$. And the extension of information-theoretic tools to the Lie-group setting in which such problems are addressed naturally is the main contribution of this paper. In Section II the mathematics of $SE(2)$ is reviewed together with stochastic nonholonomic models developed by the author and coworkers in prior work. These models will be used to motivate the new probabilistic and information-theoretic search strategies presented in Section III. Because these strategies view data obtained by a robot, as well as the motion capabilities of a robot, as issues in the differential geometry of $SE(2)$, and since "information theory on Lie groups" has not been developed in the literature, Section IV is devoted to stating some theorems that will allow methods developed by others in the context of $\mathbb{R}^n$ to be adapted easily to the Lie group setting.

A. A Review of Probabilistic Search Strategies in Euclidean Space

1) Gradient-Following: It is believed that bacteria such as E. coli obey simple rules for following resources [18], [2], [1]. In particular, they execute stochastic search strategies in $SE(3)$ whereby they swim approximately straight forward for some duration, randomize their orientation, and swim straight again. The duration of the straight segments of these trajectories is related to the strength of a chemical signal (which for lack of a better word will be called odor here); the stronger the odor, the shorter the straight portion of the path. This stochastic search approach appears to serve E.

coli well. However, at the macro scale air flows at Reynolds numbers much larger than zero, and so gradient tracking can be problematic. For this reason, a number of researchers have developed information-based methods, as reviewed in the following subsections

2) Motions Guided by the Cramér-Rao Bound: In a series of recent papers Tzanos and Zefran [22], [23], [24] develop an odor-source-location-detection algorithm based on the Cramér-Rao bound. In that approach it is assumed that a holonomic point robot takes measurements at known global positions and compares them to what the corresponding measurements would be in a database of maps in which scent fields corresponding to sources at a variety of different locations are calculated a priori. Then the Cramér-Rao bound is used to estimate which of the stored scent maps most accurately reflects the measurements that have been taken. In that formulation, maps are parameterized by source location and global knowledge of robot location. In separate work unrelated to odor plumes, Censi [5], [4] bounded the achievable accuracy of localization when a robot that does not know its position uses noisy range-finders to attempt to localize. In separate work, Cortez, Tanner, and Lumia [7] use the information-theoretic concept of relative entropy to drive mobile robots to perform radiation mapping. While these are all significant advances, none use the nonholonomic constraints inherent in the mobile robot, nor do they exploit the geometric structure of the configuration space.

3) Infotaxis: Searching for Maximum Information Gain: A probabilistic model of plume detection has been developed recently in which the goal is to determine the location of the source of an odor (e.g., smoke from a fire, chemical pathogens in the environment, etc., which are collectively referred to as "odor.") under the conditions that the density of the odor is governed by advection as well as diffusion. Under this model, a closed-form expression for the probability of registering the existence of an odor when positioned at $\mathbf{x}$ and the source is at $\mathbf{s}$ is given and denoted as $r(\mathbf{x}\,|\,\mathbf{s})$. In that work the model for the probability of the source location, $\mathbf{s}$, is given by the posterior probability density [25]

$$p_t(\mathbf{s}) = \mathcal{L}_{\mathbf{s}}(\mathcal{T}_t) / \int \mathcal{L}_{\mathbf{x}}(\mathcal{T}_t) d\mathbf{x}$$

where

$$\mathcal{L}_{\mathbf{s}}(\mathcal{T}_t) = \exp\left[-\int_0^t r(\mathbf{x}(\tau)\,|\,\mathbf{s})d\tau\right] \prod_{k=1}^{N} r(\mathbf{x}(t_k);\mathbf{s})$$

is a kind of likelihood function for the source location. Those authors then proceed using an information criterion to guide rectilinear (translation-only) motions that converge to the source location.

It should be noted that a nonholonomic robot may not be able to follow such a trajectory, and certainly not if there is noise/slip in its locomotion system. Therefore this provides an opportunity for applying Lie-theoretic tools.

As far as odor detecting robots are concerned, if two chemical sensors are located at the ends of long booms deployed at antipodal ends of a mobile robot, then this gives the robot the ability to detect gradients in its body-fixed frame. Such measurements are then naturally pose measurements. Combining this information with stochastic models of range sensing and locomotion provides a method for estimating the source location of a plume as it appears in the body-fixed frame of the robot, without global knowledge.

The literature on SLAM and odor detection have extensively used techniques from probability and information theory in Euclidean space. Concepts such as Shannon entropy, Kullback-Leibler divergence, the Cramér-Rao bound, and various filtering algorithms for data in $\mathbb{R}^n$ have been used. However, when it comes to combining stochastic models of sensor information and locomotion, there is no natural way to decouple the group-theoretic nature of the problem, unless one artificially restricts attention to holonomic robots with perfect locomotion capabilities. Unfortunately, the probablistic/information-theoretic techniques mentioned above have not been generalized to the Lie-group setting which is natural for use in robotics. However, some of the pieces are in place. For example, stochastic nonholonomic models have been developed by the author and collaborators in the context of wheeled vehicles and needle steering [27], [16] together with "kinematic state estimation" [17]. Park and coworkers have developed particle filters for $SE(n)$ [11], etc.

The main emphasis of the current paper is the development of concepts of Shannon entropy, Kullback-Leibler divergence, and the Cramér-Rao bound for pose data, and presents some new theorems about their properties.

II. METHODS

A. General Terminology

Three matrix Lie groups will play important roles in this paper:

$$SO(n) = \{R \in \mathbb{R}^{n\times n}\,|\,RR^T = \mathbb{I}, \det R = +1\};$$

$$SE(n) = \left\{ \left(\begin{array}{cc} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{array} \right) \middle| \, (R,\mathbf{t}) \in SO(n)\times\mathbb{R}^n \right\}$$

(which is equivalent to the set of rotation-translation pairs $(R,\mathbf{t})$ with the operation $\circ$ defined by $(R_1,\mathbf{t}_1)\circ(R_2,\mathbf{t}_2) = (R_1R_2, R_1\mathbf{t}_2 + \mathbf{t}_1)$); and

$$GL^+(n,\mathbb{R}) = \{A \in \mathbb{R}^{n\times n}\,|\,\det A > 0\}.$$

The group operation for each can be viewed as matrix multiplication.

From the polar decomposition of symmetric positive definite real matrices, we have that any $A \in GL^+(n,\mathbb{R})$ can be written as $A = RS$ where $S = S^T = (A^TA)^{\frac{1}{2}} \in SO(n)\backslash GL^+(n,\mathbb{R})$ and $R = A(A^TA)^{-\frac{1}{2}} \in SO(n)$. Here $SO(n)\backslash GL^+(n,\mathbb{R})$ is the homogeneous space consisting of all symmetric positive definite matrices. If $\times$ denotes the Cartesian product of spaces, then the notation $(SO(n) \times (SO(n)\backslash GL^+(n,\mathbb{R})) \rightarrow GL^+(n,\mathbb{R})$ denotes the mapping defined by the matrix multiplications $R(R^{-1}A) = A$.

Let $\{f(\mathbf{x};\boldsymbol{\theta})\,|\,\boldsymbol{\theta} \in \Theta\}$ denote a parameterized family of multivariate probability distributions. Perhaps the most common such family is the the Gaussian distribution on $\mathbb{R}^n$, for which $\boldsymbol{\theta} = (\boldsymbol{\mu}, \Sigma)$ and $\Theta = \mathbb{R}^n \times (SO(n)\backslash GL^+(n,\mathbb{R}))$. Given such a family as a model for some probabilistic phenomenon for which $\boldsymbol{\theta}$ is a priori unknown, then an estimate $\hat{\boldsymbol{\theta}} \in \Theta$ can be obtained from a set of observations $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ using an estimator. One of the most popular estimators is the method of maximum likelihood, which seeks $\hat{\boldsymbol{\theta}} \in \Theta$ such that the likelihood function is maximized. That is, $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta})$ where $L(\boldsymbol{\theta}) = \prod_{k=1}^{N} f(\mathbf{x}_k; \boldsymbol{\theta})$. This is particularly convenient when f is a Gaussian, but the concept applies equally well to other parametric families. The Cramér-Rao bound provides a limit on how accurate an estimator can be. This is not limited to the maximum likelihood estimator, which is discussed above only for the sake of concreteness.

B. Stochastic Models of Mobile Robots

The cart-like robot shown in Figure 1 moves around in the plane by turning each of its two wheels. Relative to a frame of reference fixed in the plane, the frame of reference fixed in the robot moves as a function of the torque inputs imparted by the motors to the wheels. This reference frame can be thought of as the time-dependent rigid-body motion

$$g(x,y,\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & x \\ \sin\theta & \cos\theta & y \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

where θ is the angle that the axle makes with the x-axis of the world frame, and x and y are the components of the translation of the center of the cart-like robot relative to the frame of reference fixed in the plane. The group law for $SE(2)$, which is equivalent to the multiplication of matrices of the form in (1) is

$$g(x_1,y_1,\theta_1) \circ g(x_2,y_2,\theta_2) =$$
$$g(x_1{+}x_2\cos\theta_1{-}y_2\sin\theta_1, y_1{+}x_2\sin\theta_1{+}y_2\cos\theta_1, \theta_1{+}\theta_2).$$

Furthermore, pure translational and rotational motions can be expressed as $e^{tX_1} = g(t,0,0)$, $e^{tX_2} = g(0,t,0)$, and $e^{tX_3} = g(0,0,t)$ where e^{tX_i} is the matrix exponential and X_1, X_2, X_3 are respectively the matrices

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}; \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Therefore we can write $g(x,y,\theta) = \exp(xX_1 + yX_2)\exp(\theta X_3)$. The logarithm map goes the opposite way and converts elements of $SE(2)$ into linear combinations of the X_i matrices (which are basis elements for the Lie algebra $se(2)$). The logarithm map does not work for some elements of $SE(2)$, but the set were it breaks down is a set of measure zero. Later in Section II-E the way that the logarithm map is used (under an integral) means that this set of measure zero is irrelevant.

If the robot's motion has been observed, then $g(t)$ is known for all times from $t = 0$ up to the present time. However, the exact location of the future location of the robot is uncertain until it actually happens since the wheels might slip. Given models describing these uncertainties, what will the most likely position and orientation of the robot be at a given future time?

Let the two wheels each have radii r, and let the distance between the wheels (called the wheelbase) be denoted as L. Imagine that the angles through which the wheels turn around their axes are governed by "stochastic differential equations" of the form

$$d\phi_1 = \omega(t)dt + \sqrt{D}dw_1 \tag{2}$$
$$d\phi_2 = \omega(t)dt + \sqrt{D}dw_2 \tag{3}$$

where dw_i each represent "uncorrelated unit white noise," D scales the strength of the noise, and $\omega(t)$ is what the time-rate-of-change of $\theta(t)$ would be if D were zero. Then a "stochastic trajectory" for $g(t)$ in (1) is defined by stochastic differential equations of the form [27]

$$\begin{pmatrix} dx \\ dy \\ d\theta \end{pmatrix} = \begin{pmatrix} r\omega\cos\theta \\ r\omega\sin\theta \\ 0 \end{pmatrix} dt + \sqrt{D} \begin{pmatrix} \frac{r}{2}\cos\theta & \frac{r}{2}\cos\theta \\ \frac{r}{2}\sin\theta & \frac{r}{2}\sin\theta \\ \frac{r}{L} & -\frac{r}{L} \end{pmatrix} \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix} \tag{4}$$

Stochastic trajectories, by definition, are not repeatable. However, if such an equation is simulated many times, each time starting from the same initial conditions (say, $x = y = \theta = 0$), then a function, $f(x,y,\theta;t)$ that records the distribution of positions and orientations of the cart at the same value of time, t, in each trajectory can be defined. As explained in detail in [27], [8], a well-developed theory for linking stochastic differential equations such as (4) to functions such as $f(x,y,\theta;t)$ exists. This theory produces a partial differential equation (called a *Fokker-Planck equation*) for $f(x,y,\theta;t)$. In the present context, this equation is of the form [27]

$$\frac{\partial f}{\partial t} = -r\omega\cos\theta\frac{\partial f}{\partial x} - r\omega\sin\theta\frac{\partial f}{\partial y} +$$
$$\frac{D}{2}\left(\frac{r^2}{2}\cos^2\theta\frac{\partial^2 f}{\partial x^2} + \frac{r^2}{2}\sin 2\theta\frac{\partial^2 f}{\partial x\partial y} + \frac{r^2}{2}\sin^2\theta\frac{\partial^2 f}{\partial y^2} + \frac{2r^2}{L^2}\frac{\partial^2 f}{\partial\theta^2}\right).$$

There is a very clean coordinate-free way of writing (4) and the above equation. Namely (4) can be written as

$$\left(g^{-1}\frac{dg}{dt}\right)^{\vee} dt = r\omega\mathbf{e}_1 dt + \frac{r\sqrt{D}}{2}\begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 2/L & -2/L \end{pmatrix} d\mathbf{w}$$

where $\vee$ is the "vee operator" defined in [15]. The coordinate-free version of the Fokker-Planck equation is given below.

C. Calculus on Euclidean Groups

Analogs of the usual partial derivatives in $\mathbb{R}^n$ can be defined in the Lie-group setting as [6]

$$\tilde{X}_i f = \left[\frac{d}{dt} f\left(g \circ e^{tX_i}\right)\right]\Big|_{t=0}, \; i = 1,2,3. \tag{5}$$

These are called Lie derivatives. The Fokker-Planck equation above can be written compactly in terms of these Lie derivatives as [27]

$$\frac{\partial f}{\partial t} = -r\omega\tilde{X}_1 f + \frac{r^2 D}{4}(\tilde{X}_1)^2 f + \frac{r^2 D}{L^2}(\tilde{X}_3)^2 f. \tag{6}$$

And while efficient techniques for solving this sort of equation exist for both the long-time and short time cases (see e.g. [27], [8]), the emphasis in the current paper is not solution techniques, but rather an assessment of how pose information can be obtained from (6) directly. The first observation that can be made along these lines is that if the robot starts from a known pose, $g_0 = g(x_0, y_0, \theta_0)$, then (6) is solved subject to the initial condition $f_{g_0}(g; 0) = \delta(g_0^{-1} \circ g) = \delta(x - x_0, y - y_0, \theta - \theta_0)$. It will often be convenient to set $g_0 = e = g(0, 0, 0)$, the identity, and use the notation $f_t(g)$ as shorthand for $f_e(g; t)$. If the robot runs for a time t_1 from the initial pose $g_0 = e$, the resulting pose distribution will be $f(g; t_1)$. We will not know the robot's pose at time t_1, but only a distribution of poses, the support of which contains the actual poses.

In analogy with the way that natural generalizations of partial derivatives exist for Lie groups as defined in (5), so too do generalizations of the concept of integration. In particular, for $SE(2)$ the natural volume element is $dg = dxdyd\theta$, and for $SO(3)$ it can be written in terms of ZXZ Euler angles as $dR = \sin\beta d\alpha d\beta d\gamma$. Both $SO(3)$ and $SE(2)$ (and $SE(3)$ as well) are examples of *unimodular* Lie groups, which means that integration of arbitrary integrable functions is invariant under shifts and inversions of the argument:

$$\int_G f(g)dg = \int_G f(g_0 \circ g)dg = \int_G f(g \circ g_0)dg = \int_G f(g^{-1})dg.$$

If the robot continues to move for an additional amount of time, t_2, then the distribution will be updated as a convolution over $G = SE(2)$ of the form

$$f_{t_1+t_2}(g) = (f_{t_1} * f_{t_2})(g) = \int_G f_{t_1}(h) f_{t_2}(h^{-1} \circ g)dh. \quad (7)$$

This sort of convolution has been used extensively by the author and coworkers to describe manipulator workspaces for more than a decade (see for example, [10]) but it applies equally well to probabilistic mobile robotics with travel time of a mobile robot replacing length along a manipulator arm (a fact that appears to be unknown to the SLAM community). Here the group operation, $\circ$, leads to the convolution operation, $*$, in analogy with the way that addition in $\mathbb{R}^n$ leads to the usual concept of convolution. It is the existence and properties of the convolution operation that will allow us to derive some inequalities that parallel those in classical information theory. However, the important difference that $\circ$ (and therefore $*$) is not commutative (since in general $g_1 \circ g_2 \neq g_2 \circ g_1$) means that some aspects of the classical information theory need to be modified.

D. Entropy and Relative Entropy on Euclidean Groups

Equipped with a method to integrate, all of the classical definitions of continuous information theory can be generalized to the group setting. Namely, the Shannon entropy and Kullback-Leibler divergence become

$$S(f) = -\int_G f(g) \log f(g) dg$$

and

$$D_{KL}(f \,\|\, \phi) = \int_G f(g) \log\left(\frac{f(g)}{\phi(g)}\right) dg$$

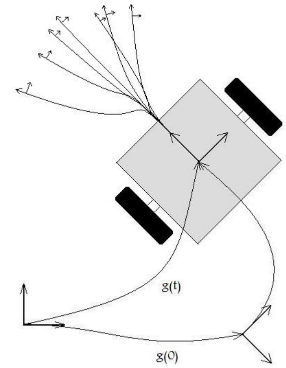

Fig. 1. A Kinematic Cart with an Uncertain Future Position and Orientation

where $f(g)$ and $\phi(g)$ are probability density functions (i.e., they are nonnegative functions that integrate to unity). Furthermore, many information inequalities formulated in Euclidean space also hold in the context of Lie groups, as exemplified by the following.

Theorem 1: The entropy of convolved pdfs increase, and the data processing inequality holds:

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

and

$$D_{KL}(f_1 \,\|\, f_2) \geq \max\left\{D_{KL}(f_1 * \phi \,\|\, f_2 * \phi), D_{KL}(\phi * f_1 \,\|\, \phi * f_2)\right\}.$$

Proof: Follows from the convexity of the functions $-\log x$ and $x \log x$, and Jensen's inequality.

E. Fisher Information and Covariance for $SE(2)$

The covariance of a concentrated probability density centered around the mean μ, which is defined as the point $\mu \in G$ such that

$$\int_G (\log g)^\vee f(\mu \circ g)dg = \mathbf{0},$$

can be defined as [26]

$$\Sigma(f) = \int_G (\log g)^\vee [(\log g)^\vee]^T f(\mu \circ g)dg. \quad (8)$$

The Fisher information matrix is defined here by its elements as

$$F_{ij}(f) = \int_G \frac{1}{f}(\tilde{X}_i f)(\tilde{X}_j f)dg, \quad (9)$$

and the scalar Fisher information is the trace of F. Note that if $\rho(g)$ is a pdf with mean at the identity and $f(g) = \rho(\mu^{-1} \circ g)$ is a shifted version with mean at μ, then

$$F(\rho) = F(f),$$

which results from the invariance of integration over G.

The author and collaborators have investigated how the covariance of the convolution of pdfs behaves, and have obtained closed-form approximations when the pdfs are "concentrated." In particular, if ρ_1 and ρ_2 are two probability distributions that are respectively centered around poses g_1

and g_2, then it can be shown that $\rho_1 * \rho_2$ is centered around $g_1 \circ g_2$ and the covariance propagates under convolution as (see [26] and references therein)

$$\Sigma_{\rho_1 * \rho_2} = Ad(g_2^{-1})\Sigma_1 Ad^T(g_2^{-1}) + \Sigma_2$$

where for $SE(2)$

$$Ad(g) = \begin{pmatrix} \cos\theta & -\sin\theta & y \\ \sin\theta & \cos\theta & -x \\ 0 & 0 & 1 \end{pmatrix}.$$

Later in the current paper the properties of Fisher information are explored and a group-theoretic version of the Cramér-Rao bound is derived (with the case of $SE(2)$ being the one of primary interest for mobile robotics).

III. PROBABILISTIC SEARCH STRATEGIES ON THE EUCLIDEAN GROUP

In this section several probabilistic strategies for localization and plume source detection by mobile robots are described as problems on $SE(2)$. It is assumed that each mobile robot has olfactory sensors at $\mathbf{r}_1, ..., \mathbf{r}_n$ and range sensors that project from the origin of the frame attached to the robot through the points $\mathbf{p}_1, ..., \mathbf{p}_n$. Both $\{\mathbf{r}_i\}$ and $\{\mathbf{p}_i\}$ are planar position vectors fixed in the body-fixed frame of the robot. The case of a sensor deployed on a movable boom or a laser range finder can be viewed within this framework by letting m and n become large to represent finely sampled data.

The problems of localization and guiding movement to the source of a plume without hitting obstacles can be implemented in a variety of ways as described in the subsections that follow.

A. Localization by Optimization on $SE(2)$

Suppose that a robot is placed at random in a known environment and it must localize itself (i.e. determine its pose relative to a world frame) rapidly using distance measurements from, e.g., a laser range finder. This will provide a set of measurements $\mu_i = d_i + n_i$ where d_i is the distance to a wall along the line segment from laser source through the point $\mathbf{p}_i$, and n_i is a corresponding noise. These measurements can be compared with the ideal values obtained from a world model interrogated with different robot poses. If $m_i(g)$ denotes the distance that the model would produce for the i^{th} sensor when the robot is assumed to be at pose g, then a measure of goodness of fit between the model and observations can be used to estimate g. For example, a two-norm of the form $c_2(g) = \sum_{i=1}^{n}(\mu_i - m_i(g))^2$ could be used, or a Kullback-Leibler-divergence-like measure of the form $c_{KL}(g) = \sum_{i=1}^{n} \mu_i(g)\log(\mu_i(g)/m_i(g))$ could be used. Regardless, the minimization of one of these costs over g would provide an estimator, which is group valued.

Normally in robot localization problems, the fact that $SE(2)$ is a group is overlooked or ignored, and standard tools from Bayesian analysis and classical information inequalities such as the Cramér-Rao bound are used (essentially treating $SE(2)$ as if $[x, y, \theta] \in \mathbb{R}^3$). This becomes problematic when factoring nonholonomic constraints and the topology of the configuration space, and becomes even more acute in three-dimensional planning problems. This is therefore one motivation for the new theorems presented in Section IV. Additional motivation is provided in the following subsection.

B. Searching Odor Plumes by Gradient Ascent on $SE(2)$

If the source of an odor plume is at the global position $\mathbf{s} \in \mathbb{R}^2$, and the resulting scent field (which for the moment is assumed to be static on the time scale of robot motion) is $\rho(\mathbf{x}; \mathbf{s})$, then the i^{th} olfactory sensor on the mobile robot will register measurement $\mu_i(g(t)) = \rho(g(t)\mathbf{r}_1\cdot; \mathbf{s}) + n_i(t)$ at time t where $n_i(t)$ is the time-varying (though stationary) noise in this sensor. By constructing a vector $\boldsymbol{\mu}(g(t)) = [\mu_1(g(t)), ..., \mu_n(g(t))]^T$ and a cost function $C(g(t)) = \frac{1}{2}\boldsymbol{\mu}(g(t))^T M \boldsymbol{\mu}(g(t))$ for some positive definite M (such as the identity matrix), then a natural way to drive the robot toward $\mathbf{s}$ (without a priori knowledge of $\mathbf{s}$) is by gradient ascent on $C(g)$. Since the robot is nonholonomic and noisy and has a non-negligible size, it is not possible to simply follow the gradient toward $\mathbf{s}$. However, we can perform a small-time average of $C(g(t))$ over the interval $t_i \leq t \leq t_i + \Delta t$ while $g(t)$ is fixed, thereby obtaining a better estimate of $C(g)$. Then a discretized version of the derivatives $\tilde{X}C$ can be computed as the robot moves.

C. Infotaxis on $SE(2)$

Two distinctly different forms of infotaxis can be perfomed on $SE(2)$ in the context of odor plume detection by mimicking the Cartesian strategies reviewed in Sections I-A.2 and I-A.3. In both forms, it will be assumed that olfactory sensors are placed far enough apart (e.g. on deployable booms) so that differences in the odor intensity distribution can be sensed. In doing so, the problem immediately becomes one that depends on the pose of the robot. The infotaxis approach in Section I-A.3 then is cast within the $SE(2)$ setting by switching trajectories $\mathbf{x}(t) \in \mathbb{R}^2$ into ones of the form $g(t) \in SE(2)$. Likewise the CRB approach reviewed in Section I-A.2, in which maps of known environments are parameterized by the global odor source location and the assumption that a robot knows its global position, can be converted to a form without global knowledge that depends on the robot pose. Specifically, if a version of the CRB for pose data existed, then the robot could, in principle, take measurements at discrete poses or along a continuous trajectory and compare these measurements with values in maps as reviewed in Section I-A.2, but since there can be uncertainty in the robot's pose, the CRB would need to be computed not only for the source position parameters that define the maps, but also the pose trajectory of the robot. In other words, the parameter in the estimation problem would become a direct product of $\mathbb{R}^2$ and one or more copies of $SE(2)$. And it would be useful to have estimation methods for this kind of group manifold.

IV. FISHER INFORMATION INEQUALITIES ON UNIMODULAR LIE GROUPS

Given the motivating problems described in the previous section, several new results are presented here that link information-theoretic inequalities to the differential-geometric properties of Lie groups that arise naturally in mobile robotics problems.

A. Fisher Information and Diffusions on Lie Groups

If $f(g,t)$ is a pdf that satisfies a diffusion equation such as (6) and its generalizations (regardless of the details of the initial conditions) then some interesting properties of $S_f(t)$ can be studied. One result is the following.

Theorem 2: If $f(g,t)$ obeys a diffusion equation on a unimodular Lie group (with or without drift) with diffusion matrix D, then the rate of entropy increase of $f(g,t)$ is related to the Fisher information of $f(g,t)$ by the equation

$$\dot{S}_f = \frac{1}{2}\mathrm{tr}[DF].$$

Proof: $\dot{S}_f = dS_f/dt$, and differentiating under the integral sign gives

$$\dot{S}_f = -\int_G \left\{ \frac{\partial f}{\partial t}\log f + \frac{\partial f}{\partial t} \right\} dg. \tag{10}$$

But from the property that a diffusion equation preserves total probability,

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

and so the second term in the braces in (10) integrates to zero.

Substitution of a diffusion equation with drift of the form

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

(which generalizes (6)) into the integral for $\dot{S}_f$ and using integration by parts gives

$$\begin{aligned}
\dot{S}_f &= -\int_G \left\{ \frac{1}{2}\sum_{i,j=1}^{n} D_{ij}\tilde{X}_i\tilde{X}_j f - \sum_{k=1}^{n} d_k\tilde{X}_k f \right\} \log f \, dg \\
&= \frac{1}{2}\sum_{i,j=1}^{n} D_{ij}\int_G \frac{1}{f}(\tilde{X}_j f)(\tilde{X}_i f)\, dg \\
&= \frac{1}{2}\mathrm{tr}[DF].
\end{aligned}$$

The integral corresponding to the drift disappears, i.e.,

$$\int_G \tilde{X}_k f \, dg = 0,$$

from the generalized Stokes theorem for manifolds [8]. Since D and F are positive semidefinite, it follows that $\dot{S}_f \geq 0$.

B. Fisher Information and Convolution on Lie 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.

The focus of this subsection is the following theorem

Theorem 3: Let $\rho_i(g)$ be smooth pdfs on the unimodular Lie group G. Then the following inequality holds for the Fisher information matrix:

$$\mathrm{tr}[F(\rho_1 * \rho_2)P] \leq \mathrm{tr}[F(\rho_i)P] \tag{11}$$

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

$$f_1(h) = \int_G f_{12}(h,g)dg = \rho_1(h)$$

and

$$f_2(g) = \int_G f_{12}(h,g)dh = (\rho_1 * \rho_2)(g).$$

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

$$\frac{(\tilde{X}_i f_2)(g)}{f_2(g)} = \int_G \frac{(\tilde{X}_i\rho_2)(k)}{\rho_2(k)}\frac{\rho_1(g\circ k^{-1})\rho_2(k)}{f_2(g)}dk. \tag{12}$$

Therefore, from the Cauchy-Schwarz inequality,

$$F_{ii}(f_2) \leq F_{ii}(\rho_2).$$

The above can be written as

$$\mathrm{tr}[F(\rho_1 * \rho_2)\Lambda] \leq \mathrm{tr}[F(\rho_2)\Lambda]$$

for any positive definite diagonal matrix Λ with the same dimensions as F. If this is true in one Lie-algebra basis $\{X_i\}$, then the more general statement in (11) must follow in another basis where $P = Q^T\Lambda Q = P^T > 0$ replaces Λ. Since the initial choice of basis is arbitrary, (11) must hold in every basis for an arbitrary positive definite matrix P. This completes the proof.

C. The Cramér-Rao Bound On Unimodular Lie Groups

The focus of this subsection is the following theorem

Theorem 4: Let $\rho(g) = \rho(g^{-1})$ be a pdf on a unimodular Lie group G (e.g. $SE(n)$ or $SO(n)$), and let $f(g;\mu) = \rho(\mu^{-1}\circ g)$. Given an unbiased estimator of μ, then the Cramér-Rao bound holds for sufficiently small $\|\Sigma\|$ in the following form

$$\Sigma \geq F^{-1} \tag{13}$$

where Σ and F are defined in (8) and (9) and the above matrix inequality is interpreted as $\lambda_i(\Sigma - F^{-1}) \geq 0$ for $i = 1, 2, ..., n$ where n is the dimension of G.

Proof: For a symmetric pdf, $\rho(g) = \rho(g^{-1})$, the mean is at the identity, i.e., $g = e$, and so

$$\int_G (\log g)^\vee \rho(g) dg = \mathbf{0}. \tag{14}$$

The invariance of integration under shifts then gives

$$\phi(\mu) = \int_G (\log(\mu^{-1} \circ g))^\vee \rho(\mu^{-1} \circ g) dg = \mathbf{0}. \tag{15}$$

Applying the derivatives $\tilde{X}_i$ to $\phi(\mu)$ gives an expression for $\tilde{X}_i \phi(\mu) = 0$ that can be expanded under the integral using the product rule $\tilde{X}_i(a \cdot b) = (\tilde{X}_i a) \cdot b + a \cdot (\tilde{X}_i b)$ where in the present case $a = (\log(\mu^{-1} \circ g))^\vee$ and $b = \rho(\mu^{-1} \circ g)$. Note that when $\rho(\cdot)$ is highly concentrated, the only values of g that significantly contribute to the integral are those for which $\mu^{-1} \circ g \approx e$. By definition

$$\begin{aligned} \tilde{X}_i (\log(\mu^{-1} \circ g))^\vee &= \frac{d}{dt} (\log((\mu \circ e^{tX_i})^{-1} \circ g))^\vee \Big|_{t=0} \\ &= \frac{d}{dt} [\log(e^{-tX_i} \circ \mu^{-1} \circ g)]^\vee \Big|_{t=0}. \end{aligned}$$

Using the Baker-Campbell-Hausdorff formula

$$\log(e^X e^Y) \approx X + Y + \frac{1}{2}[X, Y]$$

with $X = -tX_i$ and $Y = \log(\mu^{-1} \circ g)$ together with the fact that $\mu^{-1} \circ g \approx e$ then gives

$$\int_G [\tilde{X}_i (\log(\mu^{-1} \circ g))^\vee] \rho(\mu^{-1} \circ g) \approx -\mathbf{e}_i. \tag{16}$$

The second term in the expansion of $\tilde{X}_i \phi(\mu)$ is

$$\int_G [\log(\mu^{-1} \circ g)]^\vee \rho(e^{-tX_i} \circ \mu^{-1} \circ g) dg \Big|_{t=0} =$$

$$\int_G [\log h]^\vee \rho(e^{-tX_i} \circ h) dh \Big|_{t=0}$$

where the change of variables $h = \mu^{-1} \circ g$ has been made. Using the symmetry of ρ gives $\rho(e^{-tX_i} \circ h) = \rho(h^{-1} \circ e^{tX_i})$, and making the change of variables $h \to k^{-1}$ then reduces this term to $\int (\log k^{-1})^\vee (\tilde{X}_i \rho)(k) dk$. Letting $i = 1, 2, 3$, and recombining all of the parts means that $\tilde{X}_i \phi(\mu) = 0$ can be written in the form $\int_G a_i(k) b_j(k) dk = \delta_{ij}$ where $a_i(k) = [\rho(k)]^{\frac{1}{2}} (\log k^{-1})^\vee \cdot \mathbf{e}_i$ and $b_j(k) = [\rho(k)]^{\frac{1}{2}} \tilde{X}_j [\log \rho(k)]$. Then, as in the proof of the classical Cramér-Rao bound, using the Cauchy-Schwarz inequality gives the result in (13).

V. CONCLUSIONS

Information-theoretic methods on Lie groups are motivated in the context of nonholonomic mobile robotics problems. The following four classical (continuous) information-theoretic inequalities were extended to the Lie group context in this paper: (1a) entropy increases under convolution of pdfs (which is equivalent to the addition/composition of random variables); (1b) the data processing lemma; (2) the relationship between the rate of increase of entropy in a diffusion process and the Fisher information; (3) the behavior of Fisher information under convolution; (4) the Cramér-Rao bound. The recognition of these properties means that information-driven methods for robot localization and odor source detection developed in the context of holonomic robots in Euclidean spaces can be extended in a natural way to nonholonomic robots moving in $SE(2)$.

REFERENCES

[1] Berg, H.C., *E. coli in Motion*, Springer, New York, 2004.

[2] Bray, D., *Cell Movements*, Garland Pubishing, Inc., New York and London, 1992.

[3] Bullo, F., Lewis, A. D., *Geometric Control of Mechanical Systems*, Springer, 2004.

[4] Censi, A., "On Achievable Accuracy for Pose Tracking," *ICRA'09*

[5] Censi, A., "On achievable accuracy for range-finder localization," *ICRA'07*

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

[7] Cortez, R.A., Tanner, H.G., Lumia, R., "Distributed Robotic Radiation Mapping," *ISER'08*

[8] Chirikjian, G.S., *Stochastic Models, Information Theory, and Lie Groups*, Birkhäuser, 2009.

[9] Durrant-Whyte, H.F., "Sensor Models and Multisensor Integration," *IJRR*, 7(6):97-113, 1988.

[10] Ebert-Uphoff, I., Chirikjian, G.S., "Inverse Kinematics of Discretely Actuated Hyper-Redundant Manipulators Using Workspace Densities," *ICRA'96*, pp. 139-145

[11] Kwon, J., Choi, M., Park, F. C., Chu, C., "Particle filtering on the Euclidean group: framework and applications, " *Robotica*, Vol. 25, pp. 725737, 2007.

[12] Makadia, A., Daniilidis, K., "Rotation estimation from spherical images," *PAMI*, 28:11701175, 2006.

[13] Malis, E., Hamel, T., Mahony, R., Morin, P., "Dynamic Estimation of Homography Transformations on the Special Linear Group for Visual Servo Control," *ICRA'09*

[14] Mourikis, A. Roumeliotis, S., "On the treatment of relative-pose measurements for mobile robot localization," *ICRA'06*

[15] Murray, R., Li, Z., Sastry, S., *A Mathematical Introduction to Robotics*, CRC Press, 1994.

[16] Park, W., Kim, J.S., Zhou, Y., Cowan, N.J., Okamura, A.M., Chirikjian, G.S., "Diffusion-based motion planning for a nonholonomic flexible needle model," *ICRA'05*

[17] 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

[18] Patlak, C.S., "A mathematical contribution to the study of orientation of organisms," *Bulletin of Mathematical Biophysics*, Vol. 15, 1953. pp. 431-476.

[19] Porat, B., Nehorai, A., "Localizing vapor-emitting sources by moving sensors," *IEEE Trans. Signal Processing*, 44 (4):10181021, April 1996.

[20] Russell, R. A.*Odour Detection by Mobile Robots*, World Scientific, Singapore, 1999.

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

[22] Tzanos, P., Zefran, M., Nehorai, A., "Information based distributed control for biochemical source detection and localization," *ICRA'05*: pp. 4457-4462

[23] Tzanos, P., Zefran, M., "Stability analysis of information based control for biochemical source localization," *ICRA'06*: pp. 3116 - 3121

[24] Tzanos, P., Zefran, M., "Locating a circular biochemical source: Modeling and control," *ICRA'07*: pp. 523-528

[25] Vergassola, M., Villermaux, E., Shraiman, B.I., " Infotaxis as a strategy for searching without gradients, " *Nature*, Vol 445, 25 January 2007, pp. 406-409.

[26] Wang, Y., Chirikjian, G.S., "Nonparametric Second-Order Theory of Error Propagation on the Euclidean Group," *IJRR*, 27(1112): 12581273, 2008.

[27] Zhou, Y., Chirikjian, G.S., "Probabilistic Models of Dead-Reckoning Error in Nonholonomic Mobile Robots," *ICRA'03*