


Rate of Entropy Production in Stochastic Mechanical Systems

Gregory S. Chirikjian 

Department of Mechanical Engineering, National University of Singapore, Singapore 117575, Singapore; mpegre@nus.edu.sg

Abstract: Entropy production in stochastic mechanical systems is examined here with strict bounds on its rate. Stochastic mechanical systems include pure diffusions in Euclidean space or on Lie groups, as well as systems evolving on phase space for which the fluctuation-dissipation theorem applies, i.e., return-to-equilibrium processes. Two separate ways for ensembles of such mechanical systems forced by noise to reach equilibrium are examined here. First, a restorative potential and damping can be applied, leading to a classical return-to-equilibrium process wherein energy taken out by damping can balance the energy going in from the noise. Second, the process evolves on a compact configuration space (such as random walks on spheres, torsion angles in chain molecules, and rotational Brownian motion) lead to long-time solutions that are constant over the configuration space, regardless of whether or not damping and random forcing balance. This is a kind of potential-free equilibrium distribution resulting from topological constraints. Inertial and noninertial (kinematic) systems are considered. These systems can consist of unconstrained particles or more complex systems with constraints, such as rigid-bodies or linkages. These more complicated systems evolve on Lie groups and model phenomena such as rotational Brownian motion and nonholonomic robotic systems. In all cases, it is shown that the rate of entropy production is closely related to the appropriate concept of Fisher information matrix of the probability density defined by the Fokker–Planck equation. Classical results from information theory are then repurposed to provide computable bounds on the rate of entropy production in stochastic mechanical systems.

Keywords: entropy; information; inequalities; Lie Group; statistical mechanics; fluctuation-dissipation theory; covariance; stochastic modeling



Citation: Chirikjian, G.S. Rate of Entropy Production in Stochastic Mechanical Systems. *Entropy* **2022**, *24*, 19. <https://doi.org/10.3390/e24010019>

Academic Editors: Thomas Dittrich and Antonio Maria Scarfone

Received: 6 November 2021

Accepted: 22 December 2021

Published: 23 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The second law of thermodynamics introduced the concept of entropy, and states that at the macroscopic scale entropy is nondecreasing for a closed isolated system, i.e., one for which no heat enters or leaves and on which no external work is performed. This law was postulated in the mid 1800s, motivated in part by the desire for efficiency gains in steam engines. A closed system could be an isolated container on a lab bench, or the whole universe, leading to the famous phrase “the universe tends toward disorder” which is a paraphrase of a statement made by Rudolf Clausius circa 1865.

Statistical mechanical arguments were developed to explain the second law at the microscopic/molecular level and were established in large part by Ludwig Boltzmann’s classical work for ideal gases involving deterministic collision models between particles. In their famous H-theorem, Boltzmann proved that:

$$\frac{dH}{dt} \leq 0$$

where

$$H(t) \doteq \int_{\mathbb{R}^3} f(\mathbf{v}, t) \log f(\mathbf{v}, t) d\mathbf{v}$$

with $f(\mathbf{v}, t)$ being the velocity distribution of gas particles in a container at a particular value of time, t , under the assumption of isotropic spatial distribution. This expression is

related to the more general Gibbs formula for entropy given later in this introduction, which differs in sign and scale. As t increases, $f(\mathbf{v}, t)$ converges to an equilibrium distribution, $f_\infty(\mathbf{v})$, known as the Maxwell–Boltzmann distribution, where H is minimized (or entropy is maximized) for a given value of temperature.

A completely different approach to Boltzmann's is the *stochastic mechanical system* in which each element in the statistical ensemble is a classical mechanical system which is forced by a combination of conservative forces, random Brownian motion and viscous damping. A stochastic mechanical system is modeled using stochastic differential equations, and ensemble behavior is described by the corresponding Fokker–Planck equation. They are not limited to particle systems, and can either be modelled as having inertia or can be purely kinematic. They can be complex molecules or even systems with nonholonomic kinematic constraints such as robots.

The main contributions of this paper are as follows:

- (1) The recognition that stochastic mechanical models can be used in place of the original deterministic collision models used to formulate statistical mechanics;
- (2) The interpretation of these stochastic models as Itô or Stratonovich is irrelevant for systems that have nonzero mass, but that these interpretations provide different results as mass becomes zero—an effect almost never discussed in statistical mechanical works;
- (3) Conditions for a stochastic mechanical system with configuration-dependent noise and damping to reach equilibrium (i.e., the time-independent probability distribution on phase space) are established, generalizing the Einstein relations and providing a statement that is new and different than those for detailed balance, and represents a new observation in fluctuation-dissipation theory;
- (4) The aforementioned stationary pdf to which a stochastic mechanical ensemble converges is in fact the Boltzmann distribution from statistical mechanics, which is obtained via a different path than it was in statistical mechanics;
- (5) Novel solutions to diffusion equations on Lie groups are provided;
- (6) Inequalities from information theory are extended beyond their original setting (e.g., to include diffusion processes on Lie groups in addition to Euclidean spaces) and used to bound the entropy and rate of entropy production in stochastic mechanical systems such as the rigid Brownian rotor and mobile robots, which are beyond the scope of classical statistical mechanics.

Mixed in with these new results is a substantial amount of review material, since it makes little sense to present intricacies about differences between Itô and Stratonovich stochastic differential equations or to talk about diffusions on Lie groups to someone familiar only with statistical mechanics or information theory. Moreover, despite the similarities in the form of entropy in these fields, the inequalities of information theory are rarely known in statistical mechanics, and the concept of Hamiltonian mechanics and phase space which are essential in statistical mechanics may be unknown to those familiar with information theory.

Stochastic mechanical models (i.e., stochastic differential equations and associated Fokker–Planck equations) are examined here with the goal of establishing strict bounds on the rate of entropy production. Individual stochastic trajectories can be generated numerically from a stochastic differential equation, but it is an ensemble of such trajectories that conveys important statistical information about the system. These ensembles can be used to generate a family of histograms indexed by time, or the associated Fokker–Planck equation can be solved to obtain the same family of probability densities. Regardless of which approach is taken, each stochastic mechanical system has an associated family of probability density functions (indexed by time) from which entropy and its rate can be computed.

A probability density function (pdf) on a measure space (X, μ) which is indexed by time is a function $f : X \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ such that:

$$\int_X f(\mathbf{x}, t) d\mu(\mathbf{x}) = 1.$$

The entropy of f at each value of time is defined as:

$$S(t) \doteq - \int_X f(\mathbf{x}, t) \log f(\mathbf{x}, t) d\mu(\mathbf{x}). \quad (1)$$

The choice of base of the logarithm amounts to a choice of measurement units for S , as a different base will result in a different scaling of $S(t)$. Throughout this article, $\log \doteq \log_e = \ln$.

When considering a statistical mechanical system, the entropy of a time-varying probability density on phase space is defined using Gibb's formula as:

$$S_B(t) \doteq -k_B \int_{\mathbf{q}} \int_{\mathbf{p}} f(\mathbf{p}, \mathbf{q}, t) \log f(\mathbf{p}, \mathbf{q}, t) d\mathbf{p} d\mathbf{q} \quad (2)$$

where $\mathbf{q} \in Q \subseteq \mathbb{R}^n$ is a set of coordinates, and $\mathbf{p} \in \mathbb{R}^n$ are the corresponding conjugate momenta. The Lebesgue measure on the $2n$ -dimensional phase space, $d\mathbf{p} d\mathbf{q} = dq_1 \dots dq_n dp_1 \dots dp_n$, is invariant under coordinate changes, though such changes do affect the bounds of integration Q , unless $Q = \mathbb{R}^n$. The inclusion of the Boltzmann constant k_B in (2) is for consistency with the statistical mechanics literature.

It should be noted that this integral is an approximation of a sum over microstates, the size of which is specified by the Planck scale. Whereas mathematically speaking continuous entropy has no lower bound, for physical systems it does because discrete entropy computed by summing over microstates is always nonnegative. As long as the features of the probability density function (pdf) are coarser than the Planck scale, then entropy differences using this continuous integral formula and the analogous entropy formula computed by summing over discrete states will be the same. That is, $S_{cont} \neq S_{disc}$, but $\Delta S_{cont} = \Delta S_{disc}$ for physical systems.

In information theory, entropy can also be over a continuous space (in which case it is called "differential" entropy) or it can be defined over a discrete space, such as a set of symbols. It is no coincidence that the same expression (without the factor of k_B) appears with the same name in information theory. As the folklore goes, the founder of information theory, Claude Shannon had not yet fixed the name 'entropy', and had a conversation with mathematician John von Neumann, who suggested [1,2]:

"You should call it entropy, for two reasons. First, your expression has been used in statistical mechanics under that name. Second, nobody really knows what entropy is, so in a debate you will always have the advantage."

These words are not exact because it was supposedly a phone conversation (if it ever happened at all) subsequently relayed through other people.

The configurational probability density is the marginal distribution:

$$f(\mathbf{q}, t) \doteq \frac{1}{\sqrt{M(\mathbf{q})}} \int_{\mathbf{p}} f(\mathbf{p}, \mathbf{q}, t) d\mathbf{p} \quad (3)$$

which is defined in this way so that:

$$\int_{\mathbf{q} \in Q} f(\mathbf{q}, t) \sqrt{M(\mathbf{q})} d\mathbf{q} = 1.$$

The corresponding configurational entropy is:

$$S_Q(t) \doteq - \int_{\mathbf{q} \in Q} f(\mathbf{q}, t) \log f(\mathbf{q}, t) \sqrt{M(\mathbf{q})} d\mathbf{q} \quad (4)$$

where $M(\mathbf{q})$ is the metric tensor for the configuration space, which is the mass matrix in the case of a mechanical system with inertia.

A unimodular Lie group is one with an integration measure that is invariant under shifts from the left and the right. In the case when \mathbf{q} parameterizes a unimodular Lie group, G , such as the group of rotations or the Euclidean motion group, then the entropy of a time-indexed pdf $f : G \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is:

$$S_G(t) \doteq - \int_G f(g, t) \log f(g, t) dg \quad (5)$$

where the Haar measure dg takes on different appearances under changes of parametrization but the value of the integral is independent of parametrization and it is invariant under shifts of the form $g \rightarrow g_0 g$ and $g \rightarrow g g_0$ for any fixed $g_0 \in G$.

The main topic of this article is to study the rate of entropy increase in stochastically forced mechanical systems. That is, for any of the entropies defined above, to calculate or bound \dot{S} . In particular, by simply moving the time derivative inside the integral and using the product rule from Calculus,

$$\dot{S}(t) = - \int_X \left\{ \frac{\partial f}{\partial t} \log f + \frac{\partial f}{\partial t} \right\} d\mu(\mathbf{x}). \quad (6)$$

However, since $f(\mathbf{x}, t)$ is a probability density function at each value of time, whose integral is equal to 1,

$$\int_X \frac{\partial f}{\partial t} d\mu(\mathbf{x}) = \frac{d}{dt} \int_X f(\mathbf{x}, t) d\mu(\mathbf{x}) = 0,$$

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

$$\dot{S}(t) = - \int_X \frac{\partial f}{\partial t} \log f d\mu(\mathbf{x}). \quad (7)$$

It is also possible to bound the value of entropy itself. The function $\Phi(x) = -\log x$ is a convex function. Consequently, Jensen's inequality gives:

$$S = \int_X f(\mathbf{x}) \Phi(f(\mathbf{x})) d\mu(\mathbf{x}) \geq \Phi(\|f\|^2)$$

or

$$S \geq -\log(\|f\|^2) \quad (8)$$

where

$$\|f\|^2 \doteq \int_X |f(\mathbf{x})|^2 d\mu(\mathbf{x}).$$

As a consequence of Parseval's inequality, (8) can then be stated in Fourier space. This is true not only for the case of Euclidean space, but for wide classes of unimodular Lie groups [3].

Moreover, $S(t)$ can be bounded from above in some contexts. For example, on Euclidean spaces it is well known that Gaussian distributions have maximum entropy over all pdfs with a given mean and covariance. Therefore, if $f(\mathbf{x}, t)$ is an arbitrary time-evolving pdf on \mathbb{R}^d with mean $\mu_f(t)$ and covariance $\Sigma_f(t)$, and if $\rho_{\mu, \Sigma}(\mathbf{x})$ denotes a Gaussian distribution with mean μ and covariance Σ , then:

$$S_f(t) \leq S_{\rho_{\mu_f, \Sigma_f}}(t). \quad (9)$$

On a compact space such as the circle or rotation group, the uniform distribution has the absolute maximum entropy of all distributions on those spaces.

2. Rate of Entropy Production for Stochastic Processes on Euclidean Space

The stochastic mechanical models addressed here are described as stochastic differential equations forced by Brownian noise (i.e., Gaussian white noise, or equivalently, increments of Wiener processes).

2.1. Review of Stochastic Differential Equations

Stochastic differential equations (SDEs) are forced by Gaussian white noises dw_i , which are increments of uncorrelated unit-strength Wiener processes. These define Brownian motion processes. That is, each $dw_i(t)$ can be viewed as an independent random draw from a one-dimensional Gaussian distribution with zero mean and unit variance. Let d denote the dimension of a Euclidean space on which the random process evolves. For systems with inertia, $d = 2n$ is phase space, and for noninertial systems $d = n$. The independent uncorrelated unit strength white noises dw_i form the components of a vector $d\mathbf{w} \in \mathbb{R}^d$. For example, given the stochastic differential equation on \mathbb{R}^d

$$d\mathbf{x} = B d\mathbf{w}$$

where $B \in \mathbb{R}^{d \times d}$ is a constant full-rank matrix, the distribution $f(\mathbf{x}, t)$ describing the ensemble of an infinite number of trajectories will satisfy:

$$\frac{\partial f}{\partial t} = \frac{1}{2} \sum_{i,j=1}^d D_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

where $D = [D_{ij}] = BB^T$. The solution of this equation subject to initial conditions $f(\mathbf{x}, 0) = \delta(\mathbf{x} - \mathbf{0})$ is the time-varying Gaussian distribution:

$$f(\mathbf{x}, t) = \frac{1}{(2\pi)^{d/2} |Dt|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{x}^T (Dt)^{-1} \mathbf{x}\right) \quad (10)$$

where $|A|$ denotes the determinant of a square matrix A . The entropy for this can be computed in closed form as: [4]

$$S(t) = \log \left\{ (2\pi e)^{d/2} |\Sigma(t)|^{1/2} \right\}. \quad (11)$$

where $\Sigma(t) = Dt$ in this context. More generally (11) can be used as the upper bound in (9) with $\Sigma = \Sigma_f$ since entropy is independent of the mean.

As an example of (8), applying it to a Gaussian distribution on \mathbb{R}^d gives:

$$S(t) \geq \log \left\{ (4\pi)^{d/2} |Dt|^{1/2} \right\}.$$

Comparing with the exact expression in (11) verifies this since $4\pi < 2\pi e$ and $\log(x)$ is a monotonically increasing function.

There are two major kinds of stochastic differential equations (SDEs), Itô and Stratonovich. Both are forced by Gaussian white noises dw_i . In the simple example above, Itô and Stratonovich interpretations lead to the same result, but in more complex cases where the coupling matrix B is configuration-dependent, the two interpretations will differ. A brief review of the main features of these two different interpretations of SDEs is given here based on the longer exposition in [5,6].

Historically, the Itô interpretation came first. If

$$dx_i(t) = a_i(x_1(t), \dots, x_d(t), t)dt + \sum_{j=1}^m B_{ij}(x_1(t), \dots, x_d(t), t) dw_j(t) \quad \text{for } i = 1, \dots, d \quad (12)$$

is an Itô SDE describing a random process on \mathbb{R}^d , where now $B \in \mathbb{R}^{d \times m}$, then the corresponding Fokker–Planck equation governing the probability density of the ensemble of states, $f(\mathbf{x}, t)$, is [5,6]

$$\frac{\partial f(\mathbf{x}, t)}{\partial t} = - \sum_{i=1}^d \frac{\partial}{\partial x_i} (a_i(\mathbf{x}, t) f(\mathbf{x}, t)) + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} \left(\sum_{k=1}^m B_{ik}(\mathbf{x}, t) B_{kj}^T(\mathbf{x}, t) f(\mathbf{x}, t) \right), \quad (13)$$

Itô SDEs are popular in mathematical statistics contexts, measurement and filtering theory, and in finance, because of the ease with which expectations can be taken, and the associated martingale properties. In engineering contexts when modeling physical processes the Stratonovich interpretation of SDEs described below is more popular because standard rules of Calculus can be used [7,8]. For this reason the Stratonovich interpretation is popular in differential geometric contexts since moving between coordinate patches involves Calculus operations. A Stratonovich SDE describing the exact same random process as the Itô SDE given above is written as

$$dx_i(t) = a_i^s(x_1(t), \dots, x_d(t), t) dt + \sum_{j=1}^m B_{ij}(x_1(t), \dots, x_d(t), t) \circ dw_j(t) \quad \text{for } i = 1, \dots, d \quad (14)$$

where \circ is used to denote the Stratonovich interpretation of the SDE, which differs from the Itô version. These interpretations are interchangeable with drift terms related as:

$$a_i(\mathbf{x}, t) = a_i^s(\mathbf{x}, t) + \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^d \frac{\partial B_{ij}}{\partial x_k} B_{kj}. \quad (15)$$

This illustrates that there is a very simple way to interconvert between Itô and Stratonovich by either adding or subtracting the last term in (15).

The Stratonovich form of the Fokker–Planck Equation (FPE) is written as [5,6]:

$$\frac{\partial f}{\partial t} = - \sum_{i=1}^d \frac{\partial}{\partial x_i} (a_i^s f) + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial}{\partial x_i} \left[\sum_{k=1}^m B_{ik} \frac{\partial}{\partial x_j} (B_{jk} f) \right] \quad (16)$$

When B is independent of the configuration variable \mathbf{x} , Itô and Stratonovich versions of the FPE are always the same, as can be seen from (15) where the discrepancy between the two versions of drift term vanishes as the partial derivatives of B with respect to \mathbf{x} vanish. This is a sufficient condition for Itô and Stratonovich SDEs to yield the same FPE, but is not a necessary one for this to be the case, as will be seen in later examples. Similar equations hold for processes evolving on manifolds and Lie groups, as will be discussed later in the paper.

The main topic addressed here is the rate at which $S(t)$ changes. This can be observed by substituting the solution of the Fokker–Planck equations into the definition of \dot{S} in (6).

2.2. Rate of Entropy Production

Using the Stratonovich form of the FPE in (16), we arrive at the following theorem:

Theorem 1. *The rate of entropy production for $f(\mathbf{x}, t)$ governed by (16) evolving freely on Euclidean space is positive and bounded from below by*

$$\dot{S} \geq \frac{1}{2} \int_{\mathbb{R}^d} \text{trace} \left[\frac{(B(\mathbf{x}, t)^T \nabla f)(B(\mathbf{x}, t)^T \nabla f)^T}{f} \right] d\mathbf{x} \quad (17)$$

when

$$\frac{\partial}{\partial x_i} \left(a_i^s - \sum_{k,j} B_{ik} \frac{\partial B_{jk}}{\partial x_j} \right) \geq 0 \quad (18)$$

with equality holding when

$$a_i^s - \sum_{k,j} B_{ik} \frac{\partial B_{jk}}{\partial x_j} = c_i \quad (19)$$

is constant for all values of i . In the case when B is a constant and $D \doteq BB^T$, then

$$\dot{S} = \frac{1}{2} \text{trace}[DF(t)]. \quad (20)$$

where

$$F(t) = \int_{\mathbb{R}^d} \frac{(\nabla f)(\nabla f)^T}{f} d\mathbf{x}$$

is the Fisher information matrix of $f(\mathbf{x}, t)$.

Proof. The $\partial f / \partial t$ term in

$$\dot{S} = - \int_{\mathbb{R}^d} \frac{\partial f}{\partial t} \log f d\mathbf{x}$$

can be expressed in terms of spatial derivatives by substituting in the FPE. This is written as two terms using integration by parts with the surface terms at infinity vanishing due to the fact that the process evolves freely and the pdf must decay to zero at infinity. First,

$$- \int_{\mathbb{R}^d} \left\{ - \sum_{i=1}^d \frac{\partial}{\partial x_i} (a_i^s f) \right\} \log f d\mathbf{x} = - \int_{\mathbb{R}^d} \sum_{i=1}^d a_i^s \frac{\partial f}{\partial x_i} d\mathbf{x}.$$

Second,

$$- \int_{\mathbb{R}^d} \left\{ \frac{1}{2} \sum_{i,j=1}^d \frac{\partial}{\partial x_i} \left[\sum_{k=1}^m B_{ik} \frac{\partial}{\partial x_j} (B_{jk} f) \right] \right\} \log f d\mathbf{x} = \frac{1}{2} \int_{\mathbb{R}^d} \frac{1}{f} \sum_{i,j=1}^d \left[\sum_{k=1}^m B_{ik} \frac{\partial}{\partial x_j} (B_{jk} f) \right] \frac{\partial f}{\partial x_i} d\mathbf{x}.$$

Expanding

$$\frac{\partial}{\partial x_j} (B_{jk} f) = \frac{\partial B_{jk}}{\partial x_j} f + B_{jk} \frac{\partial f}{\partial x_j}$$

and recollecting terms gives

$$\dot{S} = \int_{\mathbb{R}^d} \sum_{i=1}^d \left\{ -a_i^s + \sum_{i,j=1}^d B_{ik} \frac{\partial B_{jk}}{\partial x_j} \right\} \frac{\partial f}{\partial x_i} d\mathbf{x} + \frac{1}{2} \int_{\mathbb{R}^d} \frac{1}{f} \sum_{k=1}^m \left(\sum_{i=1}^d B_{ik} \frac{\partial f}{\partial x_i} \right) \left(\sum_{j=1}^d B_{jk} \frac{\partial f}{\partial x_j} \right) d\mathbf{x}.$$

The second integral is always nonnegative, as it is a positive semi-definite quadratic form and can be written as:

$$\frac{1}{2} \text{trace} \left[\int_{\mathbb{R}^d} \frac{(B(\mathbf{x}, t)^T \nabla f)(B(\mathbf{x}, t)^T \nabla f)^T}{f} d\mathbf{x} \right].$$

The first term can have either sign. However, if

$$a_i^s - \sum_{k,j} B_{ik} \frac{\partial B_{jk}}{\partial x_j} = c_i$$

is constant, then the first integral will vanish, and if by integration by parts the derivative in the first term is transferred over, the condition in the statement of the theorem will result since $f \geq 0$. \square

When (18) holds it is clear that a looser lower bound than (17) akin to (20) can be obtained as:

$$\dot{S} \geq \frac{1}{2} \text{trace}[D_0(t)F(t)] \quad (21)$$

by constructing positive definite matrix $D_0(t) = D_0^T(t)$ such that the following matrix inequality:

$$D_0(t) \leq B(\mathbf{x}, t)B^T(\mathbf{x}, t) \quad \forall \mathbf{x} \in \mathbb{R}^d$$

is satisfied. The reason why (21) then holds is because the trace is linear and the trace of the product of positive semi-definite matrices is nonnegative, and both F and $BB^T - D_0$ are positive semi-definite.

The condition (19) and result (20) are for diffusion processes with constant diffusion tensor and drift. Given initial conditions $f(\mathbf{x}, 0) = f_0(\mathbf{x})$, the solution $f(\mathbf{x}, t)$ in this case will be of the form:

$$f(\mathbf{x}, t) = (f_0 * \rho_{t\mathbf{a}, tBB^T})(\mathbf{x}) \quad (22)$$

where $\rho_{\mu, \Sigma}(\mathbf{x})$ is a multivariate Gaussian with mean μ and covariance Σ . The convolution of any two functions $f_1, f_2 \in (L^1 \cap L^2)(\mathbb{R}^d)$ is defined as:

$$(f_1 * f_2)(\mathbf{x}) \doteq \int_{\mathbb{R}^d} f_1(\mathbf{y})f_2(\mathbf{x} - \mathbf{y}) d\mathbf{y}. \quad (23)$$

Fisher information plays an important part in probability theory [9,10] and its connections to physics also have been recognized [11–13]. For a recent review of its properties see [14].

Several inequalities from Information Theory can then be used to bound both entropy and entropy rate by quantities that are easily computable. For example, it is known that [15–17]:

$$\frac{1}{\text{tr}[F(f_1 * f_2)P]} \geq \frac{1}{\text{tr}[F(f_1)P]} + \frac{1}{\text{tr}[F(f_2)P]} \quad (24)$$

where P is any real positive definite symmetric matrix with the same dimensions as F . When $P = D$ this can then be used to give a lower bound on $\text{tr}[FD]$, and hence on \dot{S} . Moreover, one reason for the significance of (24) in information theory is that it provides a path for proving the entropy power inequality [15,16] described below.

The entropy power of a pdf $f(\mathbf{x})$ on \mathbb{R}^d was defined by Shannon as [4,18]:

$$N(f) \doteq \frac{\exp(2S(f)/d)}{2\pi e}$$

where $S(f)$ denotes the entropy of f . The entropy power inequality is:

$$N(f_1 * f_2) \geq N(f_1) + N(f_2). \quad (25)$$

Since the logarithm is a strictly increasing function, this provides a lower bound on $S(f_1 * f_2)$ and hence can be used to bound the entropy of $f(\mathbf{x}, t)$ of the form in (22). In Section 3 lower bounds on \dot{S} will be derived.

It should be noted that (24) and (25) only apply for convolution on Euclidean spaces, and do not even apply for diffusion processes on the circle. In contrast, other bounds on non-Euclidean spaces and for processes that are not necessarily homogenous diffusions are presented later in this paper.

2.3. Examples

2.3.1. Brownian Motion in Euclidean Spaces

Brownian motion in d -dimensional Euclidean space with Dirac delta initial conditions was already reviewed, with pdf (10) and entropy (11). From this, the rate of entropy production can be computed explicitly as:

$$\dot{S} = \frac{(d/2)(2\pi e)^{d/2}|D|^{1/2}t^{d/2-1}}{(2\pi e)^{d/2}|D|^{1/2}t^{d/2}} = \frac{d}{2t}.$$

The version of the Fisher information matrix in the theorem for a Gaussian is the inverse of the covariance, and hence:

$$F = (Dt)^{-1} = t^{-1}D^{-1}.$$

Consequently, (20) gives:

$$\dot{S} = \frac{1}{2} \text{trace}(t^{-1}\mathbb{I}) = \frac{d}{2t}.$$

2.3.2. Brownian Motion on the Torus/Circle

The stochastic differential equation

$$d\theta = \sqrt{D}dw$$

with constant scalar D describing Brownian motion on the unit circle has an associated Fokker–Planck equation

$$\frac{\partial f}{\partial t} = \frac{1}{2}D \frac{\partial^2 f}{\partial \theta^2}$$

which is the same as the case on the line. However, the boundary condition $\theta(\pi) = \theta(-\pi)$ is imposed rather than free boundary conditions.

Let:

$$\rho(x, t) \doteq \frac{1}{\sqrt{2\pi Dt}} e^{-\frac{x^2}{2Dt}},$$

which is the solution of the heat equation on the real line subject to initial conditions $\rho(x, 0) = \delta(x)$.

The solution of the heat equation on the circle subject to initial condition $f(\theta, 0) = \delta(\theta)$ is then [3,5,6]

$$f(\theta, t) = \sum_{k=-\infty}^{\infty} \rho(\theta - 2\pi k; t) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} e^{-Dtn^2/2} \cos n\theta \quad (26)$$

The above two equalities represent two very different ways of describing the solution. In the first equality, the Gaussian solution presented in the previous section is wrapped (or folded) around the circle. The second solution is a Fourier series solution. The first is efficient when kt is small. In such cases only $k = 0$ may be sufficient as an approximation. When kt is large, truncating $n = 1$ in the Fourier expansion may be sufficient.

Statistical quantities such as the mean and variance can be computed in closed form as:

$$\mu(t) = \int_{-\pi}^{\pi} \theta f(\theta, t) d\theta = 0$$

and

$$\sigma^2(t) = \int_{-\pi}^{\pi} \theta^2 f(\theta, t) d\theta = \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} e^{-Dtn^2/2}.$$

In both expressions in (26) summations are present which makes the exact analytical computation of logarithms, and hence entropy, problematic. However, Since the function $\Phi(x) = -\log x$ is a monotonically decreasing function, then when $a, b > 0$ we have $\Phi(a+b) < \Phi(a)$. Consequently, since $\rho > 0$ everywhere,

$$S(t) \leq - \int_{-\pi}^{\pi} f(\theta, t) \log \rho(\theta, t) d\theta = \frac{1}{2} \log(2\pi Dt) + \frac{\sigma^2(t)}{2Dt}. \quad (27)$$

In the extreme case when the distribution reaches equilibrium $f_{\infty}(\theta) = 1/(2\pi)$, this is the maximum entropy possible, and hence:

$$S(t) \leq S_{\infty} = \log(2\pi). \quad (28)$$

It should be noted that all calculations here are done relative to the measure $d\theta$. For a compact space like the circle, it is common to normalize the measure so that:

$$V = \int_V 1 dV = 1.$$

In doing so for the heat kernel on the circle this would involve redefining $dV = d\theta/2\pi$ and $f'(\theta, t) \doteq 2\pi f(\theta, t)$. This has no effect on mean and covariance, but the value of entropy is shifted such that the entropy of the uniform distribution is equal to zero and the entropy of all other distributions are negative. Rewriting (28) in a way that does not depend on the choice of normalization of measure is:

$$S(t) \leq \log V(\mathbb{S}^1). \quad (29)$$

In other words, the value of entropy not only depends on the base of the logarithm, but also on the way that the integration measure is scaled.

2.3.3. Concentration of Species Transport in Inhomogeneous Compressible Flow

The concentration $c(x, t)$ of a species in inhomogeneous compressible flow can be modeled by the Equation [19–21]:

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \left[D(x, t) \frac{\partial c}{\partial x} \right] - \frac{\partial}{\partial x} (u(x, t) c) \quad (30)$$

where $D(x, t) = D_0(1 + \kappa_0 x)^2$ and $u(x, t) = u_0(1 + \kappa_0 x)$. This is a FPE, and it is possible to work backwards to find the corresponding SDE.

We see immediately from the above partial differential Equation (pde) that if $c(x, 0)$ is normalized to be a probability density function, then $c(x, t)$ will preserve this property because integrating both sides over x gives that the right side is zero, and the time derivative on the left side commutes with the integral over x .

Equation (30) represents a one-dimensional example of what was presented in the theorem, and the rate of entropy increase is:

$$\dot{S} = \int_{-\infty}^{\infty} \frac{1}{c} D(x, t) \left(\frac{\partial c}{\partial x} \right)^2 dx + u_0 \kappa_0$$

since the drift term in the entropy rate computation simplifies as:

$$- \int_{-\infty}^{\infty} u \frac{\partial c}{\partial x} dx = \int_{-\infty}^{\infty} c \frac{\partial u}{\partial x} dx = u_0 \kappa_0 \int_{-\infty}^{\infty} c(x, t) dx = u_0 \kappa_0.$$

2.3.4. Homogeneous Transport in Couette Flow

As another example from classical fluid mechanics, consider 2D homogeneous transport in Couette flow governed by Equation [22,23]:

$$\frac{\partial c}{\partial t} = D_0 \left[\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} \right] - \frac{U_0}{H} y \frac{\partial c}{\partial x}$$

where $y \in [0, H]$ and again $c(\mathbf{x}, 0)$ is normalized to be a probability density function, and hence $c(\mathbf{x}, t)$ retains this property. In this case the region over which the equation holds is an infinite slab with the concentration and its gradient vanishing at $y = 0$ and $y = H$.

We then arrive at:

$$\dot{S} = D_0 \int_0^H \int_{-\infty}^{\infty} \frac{1}{c} \left[\left(\frac{\partial c}{\partial x} \right)^2 + \left(\frac{\partial c}{\partial y} \right)^2 \right] dx dy + \mu \cdot \mathbf{e}_2$$

where

$$\mu = \int_0^H \int_{-\infty}^{\infty} \mathbf{x} c(\mathbf{x}, t) d\mathbf{x}$$

is the mean of $c(\mathbf{x}, t)$.

The pdf for concentration in both of the above examples can be solved in closed form using the methods in [23] in which inhomogeneous processes on Euclidean space are recast as homogeneous processes on an appropriately chosen Lie group. However, the purpose of these examples is to illustrate the relationship between entropy rate and Fisher information. In the next section, a classical result of information theory is used to bound the Fisher information with covariance. Covariance can be propagated without explicit knowledge of the pdf, which will be demonstrated.

3. Bounding Rate of Entropy Production with Covariance

The version of the Fisher information matrix that appears in entropy rate computations in general is not easy to compute. An exception to this statement is when the pdf is a Gaussian and

$$F_{\text{gauss}} = \Sigma_{\text{gauss}}^{-1}.$$

For other exponential families closed form expressions are also possible. However, in general computing the time-varying Fisher information matrix for a pdf satisfying a Fokker–Planck equation will not be easy to compute.

However, it is possible to bound the Fisher information matrix with the covariance of the pdf, and covariance can be propagated as an ordinary differential equation even when no explicit solution for the time-varying pdf is known.

In this section, the Cramér–Rao Bound [24,25], a famous inequality in Euclidean statistics and information theory, is reviewed. Recently it has been extended to manifolds and Lie groups [26–31]. A second kind of inequality for compact spaces such as tori, spheres, and rotation groups was introduced in [32]. Then examples where covariance is propagated directly from the FPE are given in which the aforementioned inequalities can be put to use in bounding the rate of entropy generation. An outline of the general procedure is given here.

Let μ and Σ respectively denote the mean and covariance of a pdf $f(\mathbf{x}; \mu, \Sigma)$. Then:

$$\mu \doteq \int_{\mathbb{R}^n} \mathbf{x} f(\mathbf{x}; \mu, \Sigma) d\mathbf{x},$$

or equivalently

$$\int_{\mathbb{R}^n} (\mathbf{x} - \mu) f(\mathbf{x}; \mu, \Sigma) d\mathbf{x} = 0, \quad (31)$$

and

$$\Sigma \doteq \int_{\mathbb{R}^n} (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T f(\mathbf{x}; \mu, \Sigma) d\mathbf{x}.$$

When presented with a FPE of the form:

$$\frac{\partial f}{\partial t} = \mathcal{D}f,$$

where \mathcal{D} is as in the right hand side of (16), an ordinary differential equation describing the evolution of $\mu(t)$ and $\Sigma(t)$ can be defined as:

$$\dot{\mu} = \int_{\mathbb{R}^n} \mathbf{x} \mathcal{D}f d\mathbf{x}$$

and

$$\dot{\Sigma} = \int_{\mathbb{R}^n} (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \mathcal{D}f d\mathbf{x}$$

where integration by parts can be used in some cases to obtain explicit closed-form expressions for the integrals on the right hand side of both equations. Once $\Sigma(t)$ is obtained, it can be used to bound entropy from above using (gaussian entropy) the rate of entropy production from below using the results given in the next section.

3.1. The Cramér-Rao Bound

The Cramér-Rao Bound (or CRB) is a way to bound the covariance of an estimated statistical quantity [24,25]. Here it will not be used in its most general form. Here it will only be used in the unbiased estimation of the mean of a probability density function on \mathbb{R}^n .

In the standard derivation of the CRB, as given in [18,33] for the case of estimation of the mean, the gradient of this expression with respect to μ is computed, where $\partial f / \partial \mu$ denotes the gradient as a column vector, and $\partial f / \partial \mu^T \doteq [\partial f / \partial \mu]^T$.

Differentiation of both sides of (31) with respect to μ^T gives:

$$\frac{\partial}{\partial \mu^T} \int_{\mathbb{R}^n} [\mathbf{x} - \mu] f d\mathbf{x} = \int_{\mathbb{R}^n} [\mathbf{x} - \mu] \frac{\partial f}{\partial \mu^T} d\mathbf{x} - \mathbb{I} = \mathbb{O}.$$

Here the derivative is taken under the integral. The product rule for differentiation then is used with the fact that f is a pdf in \mathbf{x} . \mathbb{O} denotes the $m \times m$ zero matrix resulting from computing $\partial \mathbf{0} / \partial \mu^T$. That is, since the zero vector is a constant quantity, each partial derivative with respect to μ_i is zero as well, resulting in an array of zero vectors, \mathbb{O} .

The above equation can be written as [18,24,25]

$$\mathbb{I} = \int_{\mathbb{R}^n} \mathbf{a}(\mathbf{x}, \mu) \mathbf{b}^T(\mathbf{x}, \mu) d\mathbf{x} \in \mathbb{R}^{p \times m} \quad (32)$$

where

$$\mathbf{a}(\mathbf{x}, \mu) = [f]^{-\frac{1}{2}} [\mathbf{x} - \mu]$$

and

$$\mathbf{b}(\mathbf{x}, \mu) = [f]^{-\frac{1}{2}} \frac{\partial f}{\partial \mu}.$$

Using the fact that $f(\mathbf{x}; \mu, \Sigma) = f(\mathbf{x} - \mu; \mathbf{0}, \Sigma)$ means that:

$$\mathbf{b}(\mathbf{x}, \mu) = -[f]^{-\frac{1}{2}} \frac{\partial f}{\partial \mathbf{x}}.$$

Then it becomes clear that:

$$F = \int_{\mathbb{R}^n} \mathbf{b}(\mathbf{x}, \mu) [\mathbf{b}(\mathbf{x}, \mu)]^T d\mathbf{x} \quad (33)$$

and

$$\Sigma = \int_{\mathbb{R}^n} \mathbf{a}(\mathbf{x}, \mu) [\mathbf{a}(\mathbf{x}, \mu)]^T d\mathbf{x}. \quad (34)$$

Following the logic in [33], two arbitrary vectors are introduced: $\mathbf{v} \in \mathbb{R}^p$ and $\mathbf{w} \in \mathbb{R}^m$. Then (32) is multiplied on the left by \mathbf{v}^T and on the right by \mathbf{w} to give:

$$\mathbf{v}^T \mathbb{I} \mathbf{w} = \int_{\mathbb{R}^n} \mathbf{v}^T \mathbf{a}(\mathbf{x}, \mu) \mathbf{b}^T(\mathbf{x}, \mu) \mathbf{w} d\mathbf{x}. \quad (35)$$

Regrouping terms in the resulting expression, squaring, and using the Cauchy–Schwarz inequality, then gives [24,25,33]:

$$\begin{aligned} \left(\int_{\mathbb{R}^n} \mathbf{v}^T (\mathbf{a} \mathbf{b}^T) \mathbf{w} d\mathbf{x} \right)^2 &= \left(\int_{\mathbb{R}^n} (\mathbf{v}^T \mathbf{a}) (\mathbf{b}^T \mathbf{w}) d\mathbf{x} \right)^2 \\ &\leq \left(\int_{\mathbb{R}^n} (\mathbf{v}^T \mathbf{a})^2 d\mathbf{x} \right) \left(\int_{\mathbb{R}^n} (\mathbf{w}^T \mathbf{b})^2 d\mathbf{x} \right) \\ &= \left(\int_{\mathbb{R}^n} \mathbf{v}^T \mathbf{a} \mathbf{a}^T \mathbf{v} d\mathbf{x} \right) \left(\int_{\mathbb{R}^n} \mathbf{w}^T \mathbf{b} \mathbf{b}^T \mathbf{w} d\mathbf{x} \right). \end{aligned}$$

From the Equations (33)–(35), this can be written as:

$$(\mathbf{v}^T \mathbb{I} \mathbf{w})^2 \leq (\mathbf{v}^T \Sigma \mathbf{v}) (\mathbf{w}^T F \mathbf{w}).$$

Making the choice of $\mathbf{w} = F^{-1} \mathbf{v}$ yields:

$$(\mathbf{v}^T F^{-1} \mathbf{v})^2 \leq (\mathbf{v}^T \Sigma \mathbf{v}) (\mathbf{v}^T F^{-1} \mathbf{v}).$$

This simplifies to:

$$\mathbf{v}^T (\Sigma - F^{-1}) \mathbf{v} \geq 0 \quad \text{for arbitrary } \mathbf{v} \in \mathbb{R}^n \quad (36)$$

Consequently, the term in parenthesis is a positive definite matrix, or as a matrix inequality [18,24,25,33]:

$$\Sigma \geq F^{-1}, \quad (37)$$

which is the famous Cramér–Rao Bound (for the special case of an unbiased estimator of the mean). This is equivalently:

$$\Sigma^{-1} \leq F. \quad (38)$$

Then, for example, in all of the equations for entropy production in time-varying pdfs on Euclidean space presented earlier, it is possible to bound from below using a cascade of inequalities such as:

$$\text{tr}[DF] \geq \lambda_{\min}(D) \text{tr}[F] \geq \lambda_{\min}(D) \text{tr}[\Sigma^{-1}].$$

3.2. An Example

Returning to the example of species transport in a compressible 1D flow outlined in Section 2.3.3, this section illustrates how the entropy rate can be bounded from below using the CRB even when a closed-form solution for the pdf is not known.

From the FPE itself, it is possible to propagate the mean and covariance. Multiplying both sides of (30) and integrating by parts gives the following ordinary differential Equation (ODE) for the mean $\mu(t)$

$$\dot{\mu} = (2d_0\kappa_0 + u_0)(1 + \kappa_0\mu)$$

subject to initial conditions, $\mu(0) = \mu_0$. This ODE can be solved in closed form for $\mu(t)$. However, even if it could not be, it could be solved by numerical integration, which is much easier than solving the FPE. Similarly, since:

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 c(x, t) dx = -\mu^2 + \int_{-\infty}^{\infty} x^2 c(x, t) dx,$$

multiplying (30) by x^2 and integrating by parts gives a way to propagate the covariance with an ODE of the form:

$$\frac{d}{dt}(\sigma^2) = F(\mu, \sigma^2)$$

which can be solved either analytically or numerically subject to initial conditions $\sigma^2(0) = \sigma_0^2$.

It is worth noting that even in cases where such propagation of moments by FPE is not possible (for example, when higher moments creep into the equations so that there is not form closure), it is still possible to numerically generate a large ensemble of sample paths from the SDE corresponding to the FPE and compute variance (or covariance in multi-dof systems). Covariance estimation is much more stable than pdf estimation, and so using the CRB as a lower bound is more reliable than directly attempting to compute entropy, entropy rate, or Fisher information when the pdf is not known explicitly.

4. Classical Statistical Mechanics as Stochastic Mechanics

Classical statistical mechanics, as developed by Boltzmann, Maxwell, and Gibbs, states that entropy increases. For an introduction to phase space and equilibrium statistical mechanics see [34]. Nonequilibrium statistical mechanics has been studied extensively over a long period of time starting with Boltzmann and summarized in a number of books including [35–39]. Important results continue to be developed in modern time, e.g., [40]. An alternative to the classical Boltzmann–Gibbs formulation is stochastic mechanics [41–44]. The difference is that in Boltzmann’s original formulation of statistical mechanics the model describing collisions between gas molecules was deterministic. At the beginning of the twentieth century Einstein’s formulation of Brownian motion also did not explicitly model random forces, though Langevin did. In all of those early works on Brownian motion there was no concept of Wiener process or Itô or Stratonovich stochastic calculus. These are mid twentieth century constructs that came after. Consequently, revisiting results in statistical mechanics using more modern stochastic modeling techniques sheds light on some old problems and provides a basis for building connections between statistical mechanics, stochastic modeling, and information theory.

Here a Hamiltonian formulation of stochastic mechanics is used. The Hamiltonian of a mechanical system is defined as the total system energy written in terms the conjugate momenta \mathbf{p} and generalized coordinates \mathbf{q} :

$$H(\mathbf{p}, \mathbf{q}) \doteq \frac{1}{2} \mathbf{p}^T M^{-1}(\mathbf{q}) \mathbf{p} + V(\mathbf{q}). \quad (39)$$

Here $M(\mathbf{q})$ is the configuration-dependent mass matrix and

$$\mathbf{p} \doteq M(\mathbf{q}) \dot{\mathbf{q}}.$$

The beauty of the Hamiltonian formulation is that the volume in phase space (i.e., the joint \mathbf{p} - \mathbf{q} space) is invariant under coordinate changes.

4.1. Properties of Phase Space

As is well known and explained in [5,6], if \mathbf{q} and \mathbf{q}' are two different sets of coordinates, then kinetic energy is expressed as:

$$T = \frac{1}{2} \dot{\mathbf{q}}'^T M'(\mathbf{q}') \dot{\mathbf{q}}' = \frac{1}{2} \dot{\mathbf{q}}^T M(\mathbf{q}) \dot{\mathbf{q}}.$$

Then with the Jacobian relating rates of change,

$$\dot{\mathbf{q}}' = J(\mathbf{q}) \dot{\mathbf{q}},$$

it is clear that:

$$M(\mathbf{q}) = J^T(\mathbf{q}) M'(\mathbf{q}'(\mathbf{q})) J(\mathbf{q}).$$

From the above, and the definition of conjugate momenta, $p_i \doteq \partial T / \partial \dot{q}_i$,

$$\mathbf{p}' = J^{-T}(\mathbf{q}) \mathbf{p}. \quad (40)$$

Therefore, the two phase spaces have volume elements that are related as:

$$d\mathbf{p}' d\mathbf{q}' = \begin{vmatrix} J^{-T}(\mathbf{q}) & \partial(J^{-T}(\mathbf{q})\mathbf{p})/\partial\mathbf{q} \\ \mathbb{O} & J(\mathbf{q}) \end{vmatrix} d\mathbf{p} d\mathbf{q}.$$

The determinant of the upper-triangular block matrix in the above equation is equal to 1, illustrating the invariance:

$$d\mathbf{p} d\mathbf{q} = d\mathbf{p}' d\mathbf{q}'. \quad (41)$$

The key to this result is how \mathbf{p} transforms in (40). A similar result holds in the Lie group setting wherein the cotangent bundle of a Lie group can be endowed with an operation making it unimodular even when the underlying group is not [45]. This is analogous to why (4) requires the metric tensor weighting and is coordinate dependent and (41) is not.

4.2. Hamilton's Equations for a System Forced by External Noise and Damping

Hamilton's equations of motion are:

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} + F_i \quad (42)$$

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}. \quad (43)$$

where F_i are generalized external forces. In the case in which the mechanical system is forced by noise and viscous damping, then after multiplication by dt , these equations of motion become:

$$dp_i = -\frac{1}{2}\mathbf{p}^T \frac{\partial M}{\partial q_i} \mathbf{p} dt - \frac{\partial V}{\partial q_i} dt - \mathbf{e}_i^T C M^{-1} \mathbf{p} dt + \mathbf{e}_i^T B d\mathbf{w} \quad (44)$$

and

$$dq_i = \mathbf{e}_i^T M^{-1} \mathbf{p} dt \quad (45)$$

where \mathbf{e}_i is the i^{th} natural unit basis vector. Note that the configuration-dependant mass matrix $M = M(\mathbf{q})$, noise matrix $B = B(\mathbf{q})$, and damping $C = C(\mathbf{q})$ appear prominently in these equations.

Equations (44) and (45) can be written together as:

$$\begin{pmatrix} d\mathbf{q} \\ d\mathbf{p} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}(\mathbf{p}, \mathbf{q}) \\ \boldsymbol{\gamma}(\mathbf{p}, \mathbf{q}) \end{pmatrix} dt + \begin{pmatrix} \mathbb{O} & \mathbb{O} \\ \mathbb{O} & B(\mathbf{q}) \end{pmatrix} \begin{pmatrix} d\mathbf{w}' \\ d\mathbf{w} \end{pmatrix}. \quad (46)$$

(Here $d\mathbf{w}'$ multiplies zeros and hence is inconsequential).

The vector-valued function $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are defined by their entries:

$$\begin{aligned} \alpha_i &\doteq \mathbf{e}_i^T M^{-1} \mathbf{p} \\ \gamma_i &\doteq -\frac{1}{2}\mathbf{p}^T \frac{\partial M}{\partial q_i} \mathbf{p} - \frac{\partial V}{\partial q_i} - \mathbf{e}_i^T C M^{-1} \mathbf{p}. \end{aligned}$$

The Fokker–Planck corresponding to (46), which together with an initial distribution $f(\mathbf{q}, \mathbf{p}, 0) = f_0(\mathbf{q}, \mathbf{p})$ defines the family of time-evolving pdfs $f(\mathbf{q}, \mathbf{p}; t)$, is

$$\frac{\partial f}{\partial t} + \sum_{i=1}^n \frac{\partial}{\partial q_i} (\alpha_i f) + \sum_{i=1}^n \frac{\partial}{\partial p_i} (\gamma_i f) - \frac{1}{2} \sum_{k=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial p_i \partial p_j} (b_{ik} b_{kj}^T f) = 0 \quad (47)$$

where $b_{ij} = \mathbf{e}_i^T B \mathbf{e}_j$ is the i, j^{th} entry of B .

Note that for any mechanical system with inertia the diffusion is the same regardless of Itô or Stratonovich interpretation, as:

$$\frac{\partial^2}{\partial p_i \partial p_j} (b_{ik} b_{kj}^T f) = \frac{\partial}{\partial p_i} \left(b_{ik} \frac{\partial}{\partial p_j} (b_{kj}^T f) \right) = b_{ik} b_{kj}^T \frac{\partial^2 f}{\partial p_i \partial p_j}.$$

That is, even though B is configuration dependent, the structure of the FPE equations in the case of mechanical systems with inertia places partial derivatives with respect to momenta in the diffusion terms, and such partial derivatives pass through the configuration-dependent B matrix. For this reason in mechanical systems with inertia, it does not matter if Itô or Stratonovich interpretations of SDEs are used. This freedom allows the modeler to take the best of both worlds. However, when approximations are made in modeling the initial equations of motion, such as assuming that the inertia is negligible, then the above Hamiltonian formulation no longer applies and one must be very careful as to the interpretation of the SDE as Itô or Stratonovich.

4.3. The Boltzmann Distribution

The Boltzmann distribution is defined as:

$$f_{\infty}(\mathbf{q}, \mathbf{p}) \doteq \frac{1}{Z} \exp(-\beta H(\mathbf{p}, \mathbf{q})) \quad (48)$$

where $\beta \doteq 1/k_B T$ with k_B denoting Boltzmann's constant and T is temperature measured in degrees Kelvin.

The *partition function* is defined as:

$$Z = \int_{\mathbf{q}} \int_{\mathbf{p}} \exp(-\beta H(\mathbf{p}, \mathbf{q})) d\mathbf{p} d\mathbf{q}. \quad (49)$$

The reason for using the subscript ∞ in defining (48) is the following theorem.

Theorem 2. If $\mathbf{q} \in \mathbb{R}^n$ globally parameterizes the configuration manifold of a mechanical system, then the solution of the Fokker–Planck Equation (47) will satisfy

$$\lim_{t \rightarrow \infty} f(\mathbf{p}, \mathbf{q}, t) = f_{\infty}(\mathbf{p}, \mathbf{q})$$

if and only if

$$C = \frac{\beta}{2} B B^T. \quad (50)$$

Proof. We begin by noting that (47) can be simplified a bit. First,

$$\frac{\partial}{\partial q_i} (\alpha_i f) = \frac{\partial \alpha_i}{\partial q_i} f + \alpha_i \frac{\partial f}{\partial q_i}$$

and

$$\frac{\partial}{\partial p_i} (\gamma_i f) = \frac{\partial \gamma_i}{\partial p_i} f + \gamma_i \frac{\partial f}{\partial p_i}.$$

It is not difficult to show that:

$$\sum_{i=1}^n \left\{ \frac{\partial \alpha_i}{\partial q_i} + \frac{\partial \gamma_i}{\partial p_i} \right\} = \text{tr}(C M^{-1}).$$

Using this, and considering the equilibrium condition when $\partial f / \partial t = 0$ then reduces (47) to

$$\text{tr}(CM^{-1})f + \sum_{i=1}^n \left\{ \alpha_i \frac{\partial f}{\partial q_i} + \gamma_i \frac{\partial f}{\partial p_i} \right\} - \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial p_i \partial p_j} ((BB^T)_{ij} f) = 0, \quad (51)$$

where the substitution

$$(BB^T)_{ij} = \sum_{k=1}^n b_{ik} b_{kj}^T$$

has been made.

Note that:

$$\begin{aligned} \frac{\partial f_\infty}{\partial q_i} &= -\beta \left(\frac{\partial V}{\partial q_i} + \frac{1}{2} \mathbf{p}^T \frac{\partial M}{\partial q_i} \mathbf{p} \right) f_\infty, \\ \frac{\partial f_\infty}{\partial p_i} &= -\beta \mathbf{e}_i^T M^{-1} \mathbf{p} f_\infty = -\beta \alpha_i f_\infty, \end{aligned}$$

and hence significant cancelation results in:

$$\sum_{i=1}^n \left\{ \alpha_i \frac{\partial f_\infty}{\partial q_i} + \gamma_i \frac{\partial f_\infty}{\partial p_i} \right\} = \beta \boldsymbol{\alpha}^T C \boldsymbol{\alpha}.$$

Moreover,

$$\frac{\partial^2 f_\infty}{\partial p_i \partial p_i} = \left(-\beta m_{ij}^{-1} + \beta^2 \alpha_i \alpha_j \right) f_\infty$$

Substituting into (51) therefore gives:

$$\text{tr}(CM^{-1}) + \beta \boldsymbol{\alpha}^T C \boldsymbol{\alpha} - \frac{\beta}{2} \text{tr}(M^{-1} BB^T) - \frac{\beta^2}{2} \boldsymbol{\alpha}^T BB^T \boldsymbol{\alpha} = 0.$$

This shows that $f_\infty(\mathbf{p}, \mathbf{q})$ is in fact a solution of (51), if (50) holds. The necessary conditions for the above equality to hold boil down to the necessary conditions for the two independent statements:

$$\text{tr} \left(\left(C - \frac{\beta}{2} BB^T \right) M^{-1} \right) = 0$$

and

$$\boldsymbol{\alpha}^T \left(C - \frac{\beta}{2} BB^T \right) \boldsymbol{\alpha} = 0$$

to hold. The independence of these follows from the fact that some terms depend on $\boldsymbol{\alpha}$ and others do not, and the main equality must hold for all values of $\boldsymbol{\alpha}$.

The only way that both of the above can hold is if $C - \frac{\beta}{2} BB^T$ is skew symmetric. However, damping matrices, like stiffness and mass matrices, are symmetric, as is BB^T . Hence (50) must hold. \square

Note that the necessary and sufficient conditions in (50) for the Boltzmann distribution to be the equilibrium/stationary solution holds even when B and C are dependant on \mathbf{q} . As such, this is a generalization of the fluctuation dissipation theorem (which is stated for particles) to the case of complex mechanical systems that can be modeled as a collection of rigid bodies (e.g., biological macromolecules).

4.4. Marginal Densities and the Conundrum as Mass Becomes Zero

Marginal densities of $f(\mathbf{p}, \mathbf{q}, t)$ can be defined as:

$$f(\mathbf{p}, t) \doteq \int_{\mathbf{q}} f(\mathbf{p}, \mathbf{q}, t) |\det M(\mathbf{q})|^{1/2} d\mathbf{q}$$

and

$$f(\mathbf{q}, t) \doteq |\det M(\mathbf{q})|^{-1/2} \int_{\mathbf{p}} f(\mathbf{p}, \mathbf{q}, t) d\mathbf{p}, \quad (52)$$

which is consistent with (3).

In the equilibrium case it is always possible to compute $f_{\infty}(\mathbf{q})$ in closed form as:

$$f_{\infty}(\mathbf{q}) = \frac{1}{Z_c} e^{-\beta V(\mathbf{q})} \quad (53)$$

where

$$Z_c = \int_{\mathbf{q}} e^{-\beta V(\mathbf{q})} |\det M(\mathbf{q})|^{1/2} d\mathbf{q}$$

is the configurational partition function. Then,

$$\int_{\mathbf{q}} f_{\infty}(\mathbf{q}) |\det M(\mathbf{q})|^{1/2} d\mathbf{q} = 1.$$

In contrast, in general $f_{\infty}(\mathbf{p})$ can only be computed easily in closed form when $M(\mathbf{q}) = M_0$ is constant. In this case:

$$f_{\infty}(\mathbf{p}) = \frac{\beta^{d/2}}{(2\pi)^{d/2} |\det M_0|^{d/2}} \exp\left(-\frac{\beta}{2} \mathbf{p}^T M_0^{-1} \mathbf{p}\right)$$

is a Gaussian distribution in the momenta.

Though (53) degenerates as the inertia goes to zero, it does so gracefully since both $|\det M|^{1/2}$ and Z_c approach zero in the same way as the system mass goes to zero. We can then use it as the baseline truth to compare approximations in which inertia is neglected. For example, consider the spring-mass-damper with noise:

$$m\ddot{x} + c(x)\dot{x} + kx = b(x)n$$

where $c(x)$ and $b(x)$ are nonlinear functions satisfying the condition $2c(x) = \beta b(x)^2$. If $ndt = dw$, then as $m \rightarrow 0$, we have a conundrum unless $c = c_0$ and $b = b_0$ are constant. Namely, which of the following interpretations is correct?

$$dx_1 = -k c(x_1)^{-1} x_1 dt + 2\beta^{-1} b(x_1)^{-1} dw$$

or

$$dx_2 = -k c(x_2)^{-1} x_2 dt + 2\beta^{-1} b(x_2)^{-1} \odot dw?$$

It did not matter when there was inertia, as both gave the same FPEs in the case, but making the approximation that the mass is zero creates a situation where a choice now must be made.

The answer can be informed by comparing the corresponding pdfs that solve the associated Fokker–Planck equations, $f_1(x, t)$ and $f_2(x, t)$, with $f(x, t)$ in (52). Short of that, we can examine the behavior of the mean as a function of time, and the behavior of the equilibrium distributions as compared with (52).

The Fokker–Planck equations corresponding to the above SDEs are, respectively:

$$\frac{\partial f_1}{\partial t} = k \frac{\partial}{\partial x} (c^{-1} x f_1) + 2\beta^{-2} \frac{\partial^2}{\partial x^2} (b^{-2} f_1)$$

and

$$\frac{\partial f_2}{\partial t} = k \frac{\partial}{\partial x} (c^{-1} x f_2) + 2\beta^{-2} \frac{\partial}{\partial x} \left(b^{-1} \frac{\partial}{\partial x} (b^{-1} f_2) \right).$$

Expanding and considering equilibrium conditions gives

$$\Delta_1 = k \frac{\partial}{\partial x} (c^{-1} x f_1) + 2\beta^{-2} \frac{\partial}{\partial x} \left(-2b^{-3} \frac{\partial b}{\partial x} f_1 + b^{-2} \frac{\partial f_1}{\partial x} \right)$$

and

$$\Delta_2 = k \frac{\partial}{\partial x} (c^{-1} x f_2) + 2\beta^{-2} \frac{\partial}{\partial x} \left(-b^{-3} \frac{\partial b}{\partial x} f_2 + b^{-2} \frac{\partial f_2}{\partial x} \right)$$

where an exact solution would give $\Delta_i = 0$.

The exact configurational marginal from the Hamiltonian formulation is

$$f_\infty(x) = \left(\frac{\beta k}{2\pi} \right)^{\frac{1}{2}} e^{-\beta k x^2 / 2},$$

and it has the property

$$\frac{\partial f_\infty}{\partial x} = -\beta k x f_\infty.$$

Substituting into the above, and observing that

$$k \frac{\partial}{\partial x} (c^{-1} x f_\infty) + 2\beta^{-2} \frac{\partial}{\partial x} \left(b^{-2} \frac{\partial f_\infty}{\partial x} \right) = 0$$

due to the relationship between b and c , then

$$\Delta_1 = 2\beta^{-2} \frac{\partial}{\partial x} \left(-b^{-3} \frac{\partial b}{\partial x} f_\infty \right) = 2\Delta_2.$$

This means that neither interpretation gives the true answer at equilibrium, but the magnitude of the discrepancy in the Stratonovich model is half that of the Itô. For this reason, unless modeling systems in phase space, or if there are physical grounds for choosing a particular SDE (e.g., working backwards from Fick's Law), it is safest to consider diffusions with constant diffusion tensors, as will be the case throughout the remainder of this paper.

5. Stochastic Systems on Unimodular Lie Groups

A stochastic mechanical system that has a Lie group as its configuration space can be studied in a coordinate-free way [5,6]. These systems can be purely kinematic, or can have inertia. Concrete examples are used here to illustrate, and then general theorems are provided to quantify the rate of entropy production. Different connections between Lie groups and thermodynamics than what is presented here have been made in the literature [46–49].

5.1. Review of Unimodular Matrix Lie Groups with $SO(3)$ and $SE(2)$ as Examples

The use of geometric (and particularly Lie-theoretic) methods in the control of mechanical systems and robots has been studied extensively over the past half century [50–54]. The material and notation in this section summarizes more in-depth treatments in [3,5,6].

A matrix Lie group is a group with elements that are matrices, for which group multiplication is matrix multiplication, and for which the underlying space is an analytic manifold with the operations of group multiplication and inversion of elements being analytic also. Intuitively, matrix Lie groups are continuous families of invertible matrices with structure that is preserved under multiplication and inversion. The dimension of a matrix Lie group is the dimension of its manifold, not the dimension of the square matrices describing its elements.

For example, the group of rigid-body displacements in the Euclidean plane, $SE(2)$, can be described with elements of the form:

$$g(x, y, \theta) = \begin{pmatrix} \cos \theta & -\sin \theta & x \\ \sin \theta & \cos \theta & y \\ 0 & 0 & 1 \end{pmatrix}. \quad (54)$$

The dimension is 3 because there are three free parameters (x, y, θ) . This group is not compact as x and y can take values on the real line.

The group of pure rotations in 3D can be described by the rotation matrices:

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

$SO(3)$ is a compact 3-dimensional manifold. Again, the fact that the dimension of the matrices is also 3 is coincidental.

A unimodular Lie group is defined by the property that a measure dg can be constructed such that the integral over the group has the property that

$$\int_G f(g) dg = \int_G f(g_0 g) dg = \int_G f(g g_0) dg \quad (55)$$

for any fixed $g_0 \in G$ and any function $f \in L^1(G)$. It can also be shown that as a consequence of (55):

$$\int_G f(g) dg = \int_G f(g^{-1}) dg. \quad (56)$$

These properties are natural generalizations of those familiar to us for functions on Euclidean space.

As we are primarily concerned with probability density functions for which

$$\int_G f(g) dg = 1,$$

these clearly meet the condition of being in $L^1(G)$.

In the case of $SO(3)$ the bi-invariant measure expressed in terms of $Z - X - Z$ Euler angles (α, β, γ) is $dR = \sin \beta d\alpha d\beta d\gamma$. In the case of $SE(2)$, the bi-invariant measure is $dg = dx dy d\theta$.

The convolution of probability density functions on a unimodular Lie group is a natural operation, and is defined as:

$$(f_1 * f_2)(g) \doteq \int_G f_1(h) f_2(h^{-1}g) dh. \quad (57)$$

The convolution of two probability density functions is again a probability density.

In addition to being natural spaces over which to integrate probability density functions, natural concepts of directional derivatives of functions exist in the matrix Lie group setting. This builds on the fact that associated with every matrix Lie group is a matrix Lie algebra.

In the case of $SO(3)$, the Lie algebra consists of 3×3 skew-symmetric matrices of the form

$$X = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix} = \sum_{i=1}^3 x_i E_i. \quad (58)$$

The matrices $\{E_i\}$ form a basis for the set of 3×3 skew-symmetric matrices. The coefficients $\{x_i\}$ are all real. The notation relating the matrix X and the vector $\mathbf{x} = [x_1, x_2, x_3]^T$ is [51,52]

$$\mathbf{x} = X^\vee \text{ and } X = \hat{\mathbf{x}}. \quad (59)$$

This is equivalent to identifying E_i^\vee with \mathbf{e}_i .

For $SE(2)$ the basis elements are different, and are of the form

$$E'_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad E'_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}; \quad E'_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Every element of the Lie algebra associated with $SE(2)$ can be written as a linear combination of these, and the notation (59) is still used to identify these matrices with natural unit basis vectors $\mathbf{e}_i \in \mathbb{R}^3$. For example, $X' = x'_1 E'_1 + x'_2 E'_2 + x'_3 E'_3$. (Here the primes are used so as not to confuse the Lie algebra elements for $SE(2)$ with those for $SO(3)$, but when working with a single Lie group and Lie algebra the primes will be dropped, as in the discussion below which is for the generic case.)

For an arbitrary unimodular matrix Lie group, a natural concept of directional derivative is

$$(\tilde{X}f)(g) \doteq \left. \frac{d}{dt} f(g \exp(tX)) \right|_{t=0}. \quad (60)$$

Here the argument of the function f is read as the product of g and $\exp(tX)$, which are each in G , as is the product. If $X = \sum_i x_i E_i$ for constants $\{x_i\}$, this derivative has the property

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

Such derivatives appear in invariant statements of Fokker–Planck equations on unimodular Lie groups. Moreover, these derivatives can be used together with integration to state results such as integration by parts

$$\int_G f_1(g) (\tilde{E}_i f_2)(g) dg = - \int_G f_2(g) (\tilde{E}_i f_1)(g) dg.$$

There are no surface terms because either group is infinite in its extent (and so the functions must decay to zero at the boundaries), or it is compact (in which case the functions must match values when arriving from different directions), or both for a group such as $SE(2)$.

5.2. The Noisy Kinematic Cart

The stochastic kinematic cart has been studied extensively in the robotics literature [5,6,55–58]. In this model (which is like a motor-driven wheelchair) the two wheels each have radii r , and the wheelbase (distance between wheels) is denoted as L . The nonholonomic equations of motion are

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} \frac{r}{2} \cos \theta & \frac{r}{2} \cos \theta \\ \frac{r}{2} \sin \theta & \frac{r}{2} \sin \theta \\ \frac{r}{L} & -\frac{r}{L} \end{pmatrix} \begin{pmatrix} \dot{\phi}_1 \\ \dot{\phi}_2 \end{pmatrix}. \quad (61)$$

When the wheel rates consist of a constant deterministic part and a stochastic part, then

$$d\phi_1 = \omega dt + \sqrt{D} dw_1 \quad (62)$$

$$d\phi_2 = \omega dt + \sqrt{D} dw_2 \quad (63)$$

and multiplying (61) by dt and substituting in (63) results in an SDE. This is an example where it does not matter whether the SDE is of Itô or Stratonovich type, even though B

is not constant. The corresponding Fokker–Planck equation for the probability density function $f(x, y, \theta; t)$ with respect to measure $dx dy d\theta$ is [58]:

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

which is subject to the initial conditions $f(x, y, \theta; 0) = \delta(x - 0)\delta(y - 0)\delta(\theta - 0)$.

The coordinates (x, y, θ) that define the position and orientation of the cart relative to the world frame are really parameterizing the group of rigid-body motions of the plane, $SE(2)$. Each element of this unimodular Lie group can be described as homogeneous transformation matrices of the form in (54) in which case the group law is matrix multiplication.

Then (61) can be written in coordinate-free notation as:

$$\left(g^{-1} \frac{dg}{dt} \right)^\vee = A \dot{\phi} \quad \text{where} \quad A = \frac{r}{2} \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 2/L & -2/L \end{pmatrix}. \quad (64)$$

Here the notation \vee is used as in [3,5,6,51] in analogy with (59), but for the case of $SE(2)$ rather than $SO(3)$.

The coordinate-free version of the above Fokker–Planck equation can be written compactly in terms of these Lie derivatives as [58]:

$$\frac{\partial f}{\partial t} = -r\omega \tilde{E}_1' f + \frac{r^2 D}{4} (\tilde{E}_1')^2 f + \frac{r^2 D}{L^2} (\tilde{E}_3')^2 f \quad (65)$$

with initial conditions $f(g; 0) = \delta(g)$. The resulting time-evolving pdf is denoted as $f(g; t)$ with respect to the natural bi-invariant integration measure for $SE(2)$, which is $dg = dx dy d\theta$. Solutions for (65) can be obtained in different regimes (small Dt and large Dt) either using Lie-group Gaussian distributions or Lie-group Fourier expansions, as in [58,59]. That is not the goal here. Instead, the purpose of this example is to provide a concrete case for the derivations that follow regarding rate of entropy production.

It should be noted that degenerate diffusions on $SE(2)$ occur not only in this problem, in models of the visual cortex [60–65]. Phase noise is a problem in coherent optical communication systems that has been identified in the literature [66–70]. The Fokker–Planck equations describing phase noise have been developed and solved using various methods [71–73]. Remarkably, these FPEs are the same kind as those for the kinematic cart, inpainting, visual cortex modeling, etc. Moreover, the natural extension of (64) and (65) to $SE(3)$ has found applications in modeling DNA (as reviewed in [3,5,6]) and flexible steerable needles for robotic surgery [57,74–76].

5.3. Rotational Brownian Motion

Starting with Perrin [77], various efforts at mathematical modeling of rotational Brownian motion has been undertaken over the past century [78–82]. These include both inertial and noninertial theories. A major application is in the spectroscopy of macromolecules [83,84]. Essentially the same mathematics is applicable to modeling the time-evolving uncertainty in mechanical gyroscopes [85].

Brownian motion on Riemannian manifolds and Lie groups also has been studied over a long period of time in the mathematics literature [86–91], with the rotation group and three-sphere being two very popular objects [92–94]. In addition to forcing by white noise, forcing by Lévy processes (white noise with jumps) has also been investigated [95].

5.3.1. Inertial Theory

Euler's equation of motion for a rotating rigid body subjected to an external potential, noise, and damping can be written as:

$$I_0 d\boldsymbol{\omega} + \boldsymbol{\omega} \times (I_0 \boldsymbol{\omega}) dt = -(\tilde{\mathbf{E}}V)(R)dt - C_0 \boldsymbol{\omega} dt + B_0 d\mathbf{w} \quad (66)$$

where

$$(\tilde{\mathbf{E}}V)(R) = \begin{pmatrix} (\tilde{E}_1 V)(R) \\ (\tilde{E}_2 V)(R) \\ (\tilde{E}_3 V)(R) \end{pmatrix}.$$

Here $\boldsymbol{\omega}$ is the body-fixed description of angular velocity, which is related to a time-evolving rotation matrix (using the hat notation in (59)) as:

$$\dot{R} = R\hat{\boldsymbol{\omega}}. \quad (67)$$

The moment of inertia matrix, I_0 , damping matrix, C_0 , and noise matrix B_0 are all constant. Equations (66) and (67) define a stochastic process evolving on the tangent bundle of $SO(3)$.

This can be re-written using angular momentum, $\boldsymbol{\pi} = I_0 \boldsymbol{\omega}$, as

$$d\boldsymbol{\pi} = \boldsymbol{\pi} \times (I_0^{-1} \boldsymbol{\pi}) dt - (\tilde{\mathbf{E}}V)(R)dt - C_0 I_0^{-1} \boldsymbol{\pi} dt + B_0 d\mathbf{w}. \quad (68)$$

Equations (68) and

$$\dot{R} = R \widehat{I_0^{-1} \boldsymbol{\pi}}$$

define a stochastic process on the cotangent bundle of $SO(3)$.

Note that $\mathbf{p} \neq \boldsymbol{\pi}$. To see this, expand angular velocity and kinetic energy in coordinates as $\boldsymbol{\omega} = J(\mathbf{q})\dot{\mathbf{q}}$ and

$$T = \frac{1}{2} \dot{\mathbf{q}}^T J(\mathbf{q})^T I_0 J(\mathbf{q}) \dot{\mathbf{q}}.$$

Consequently $M(\mathbf{q}) = J(\mathbf{q})^T I_0 J(\mathbf{q})$ and $\mathbf{p} = M(\mathbf{q})\dot{\mathbf{q}}$. In contrast, $\boldsymbol{\pi} = I_0 J(\mathbf{q})\dot{\mathbf{q}}$. Therefore, in order to use the general results from statistical mechanics, the interconversion

$$\mathbf{p} = J(\mathbf{q})^T \boldsymbol{\pi}$$

must be done. Moreover, C_0 in the above equations is not the same as C in the Hamiltonian formulation. A Rayleigh dissipation function will be of the form

$$\mathcal{R} = \frac{1}{2} \boldsymbol{\omega}^T C_0 \boldsymbol{\omega} = \frac{1}{2} \dot{\mathbf{q}}^T J(\mathbf{q})^T C_0 J(\mathbf{q}) \dot{\mathbf{q}},$$

indicating that $C(\mathbf{q}) = J(\mathbf{q})^T C_0 J(\mathbf{q})$. Then converting (68) to the Hamiltonian form, the viscous and noise terms become:

$$J(\mathbf{q})^T [-C_0 I_0^{-1} \boldsymbol{\pi} dt + B_0 d\mathbf{w}] = -J(\mathbf{q})^T C_0 I_0^{-1} I_0 J(\mathbf{q}) d\mathbf{q} + J(\mathbf{q})^T B_0 d\mathbf{w}.$$

If $C(\mathbf{q}) = J(\mathbf{q})^T C_0 J(\mathbf{q})$ and $B(\mathbf{q}) = J(\mathbf{q})^T B_0$, and if J is invertible, then the condition in (50) then becomes completely equivalent to:

$$C_0 = \frac{\beta}{2} B_0 B_0^T \quad (69)$$

The structure of the C_0 matrix for a rigid body is a function of its shape. For example, the viscous drag on an ellipsoid was characterized in [96]. Given C_0 , it is possible to define $B_0 = C_0^{1/2}$.

When $I = \mathbb{I}$ and $V = 0$ (68) becomes

$$d\boldsymbol{\pi} = -\frac{\beta}{2} B_0 B_0^T \boldsymbol{\pi} dt + B_0 d\mathbf{w}. \quad (70)$$

This is an Ornstein-Uhlenbeck process, and the corresponding Fokker-Planck equation can be solved for $f(\boldsymbol{\pi}, t)$ in closed form as a time-varying Gaussian if the initial conditions are $f(\boldsymbol{\pi}, 0) = \delta(\boldsymbol{\pi})$. The equilibrium solution is the Boltzmann distribution

$$f_{\infty}(\boldsymbol{\pi}) = c(\beta) \exp\left(-\frac{\beta}{2} \|\boldsymbol{\pi}\|^2\right)$$

where $c(\beta)$ is the usual normalizing constant for a Gaussian distribution.

5.3.2. Noninertial Theory

When the inertia is negligible, as it is in the case of rotational Brownian motion of molecules, then (69) and (66) give

$$\boldsymbol{\omega} dt = B_1 d\mathbf{w} \text{ where } B_1 = \frac{2}{\beta} B_0^{-T} \dots \quad (71)$$

This can be expressed in coordinates as a Stratonovich equation:

$$\dot{\mathbf{q}} = J^{-1}(\mathbf{q}) B_1 \mathbf{w}, \quad (72)$$

or it can be kept in the invariant form (71). The corresponding Fokker-Planck equation is of the form:

$$\frac{\partial f}{\partial t} = \sum_{i,j=1}^3 D_{ij} \tilde{E}_i \tilde{E}_j f$$

where $D = B_1 B_1^T$ and each \tilde{E}_i is as in (60) with $X = E_i$.

The short-time solution of this equation subject to Dirac delta initial conditions is the Gaussian in exponential coordinates. Hence, for short times, entropy and entropy rate can be computed in closed form using the results from the Euclidean case.

In noninertial theory a special case is isotropic diffusions. Let:

$$\nabla^2 \doteq \tilde{E}_1^2 + \tilde{E}_2^2 + \tilde{E}_3^2.$$

An isotropic driftless diffusion on $SO(3)$ is one of the form

$$\frac{\partial f}{\partial t} = K \nabla^2 f. \quad (73)$$

The heat kernel for $SO(3)$ is the solution of this subject to the initial condition $f(R, 0) = \delta(R)$.

Rotation matrices can be expressed in terms of the axis and angle of rotation using Euler's formula:

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

where $\hat{\mathbf{n}}(\nu, \lambda)$ is a skew symmetric matrix such that for arbitrary vector $\mathbf{v} \in \mathbb{R}^3$ and vector cross product \times ,

$$\hat{\mathbf{n}}(\nu, \lambda) \mathbf{v} = \mathbf{n}(\nu, \lambda) \times \mathbf{v}$$

and

$$\mathbf{n}(\nu, \lambda) = \begin{pmatrix} \sin \nu \cos \lambda \\ \sin \nu \sin \lambda \\ \cos \nu \end{pmatrix}.$$

There are several different ways to choose the ranges of these coordinates to fully parameterize $SO(3)$. One way is to view these coordinates as a solid ball of radius π in which $\theta \in [0, \pi]$ serves as the radius and $\nu \in [0, \pi]$ and $\lambda \in [0, 2\pi)$, are the usual spherical angles. Another way is to let $\theta \in [0, 2\pi)$ and cut the range of one of the other variables in half. For example, $\nu \in [0, \pi/2]$ and $\lambda \in [0, 2\pi)$ would restrict \mathbf{n} to the upper hemisphere and $\nu \in [0, \pi]$ and $\lambda \in [0, \pi)$ would be like the western hemisphere (if the initial datum is chosen appropriately). In these hemispherical boundary models, the great circle that bounds the hemisphere will be half open and half closed so as not to redundantly parameterize.

There are benefits to each of these parameterizations. For example, allowing the $[0, 2\pi)$ range for θ reflects that for fixed \mathbf{n} rotations around a fixed axis bring back to the same location. That is the ‘little group’ of rotations around \mathbf{n} isomorphic to $SO(2)$ is the ‘maximal torus’ in $SO(3)$. Likewise parameterizing the whole sphere has value. For this reason, the best of both worlds can be achieved by double covering rotations by allowing both ranges to be expanded. Moreover, each range $[0, 2\pi)$ can be replaced with $[-\pi, \pi)$. Then when performing integration all that needs to be done is to divide by 2 afterwards.

Using these parameters, the integration measure dR such that the volume of $SO(3)$ is normalized to 1 is

$$dR = \frac{1}{4\pi^2} \sin^2(\theta/2) \sin \nu \, d\theta d\lambda d\nu. \quad (74)$$

When computing integrals,

$$\begin{aligned} \int_{SO(3)} f(R) \, dR &= \frac{1}{2\pi^2} \int_{\nu=0}^{\pi} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} f(R(\theta, \nu, \lambda)) \sin^2(\theta/2) \sin \nu \, d\theta d\lambda d\nu \\ &= \frac{1}{2\pi^2} \int_{\nu=0}^{\pi/2} \int_{\lambda=0}^{2\pi} \int_{\theta=-\pi}^{\pi} f(R(\theta, \nu, \lambda)) \sin^2(\theta/2) \sin \nu \, d\theta d\lambda d\nu \\ &= \frac{1}{2\pi^2} \int_{\nu=0}^{\pi} \int_{\lambda=0}^{\pi} \int_{\theta=-\pi}^{\pi} f(R(\theta, \nu, \lambda)) \sin^2(\theta/2) \sin \nu \, d\theta d\lambda d\nu \end{aligned}$$

Doubling the range gives

$$\int_{SO(3)} f(R) \, dR = \frac{1}{4\pi^2} \int_{\nu=0}^{\pi} \int_{\lambda=0}^{2\pi} \int_{\theta=-\pi}^{\pi} f(R(\theta, \nu, \lambda)) \sin^2(\theta/2) \sin \nu \, d\theta d\lambda d\nu \quad (75)$$

and when $f(R(\theta, \nu, \lambda)) = f(\theta) = f(-\theta)$

$$\begin{aligned} \int_{SO(3)} f(R) \, dR &= \frac{2}{\pi} \int_{\theta=0}^{\pi} f(\theta) \sin^2(\theta/2) \, d\theta \\ &= \frac{1}{\pi} \int_{\theta=-\pi}^{\pi} f(\theta) \sin^2(\theta/2) \, d\theta. \end{aligned} \quad (76)$$

All normalizations are such that:

$$\int_{SO(3)} 1 \, dR = 1.$$

The Laplacian operator for $SO(3)$ in this axis-angle parameterization is [3,97,98]

$$\nabla^2 = \frac{\partial^2}{\partial \theta^2} + \cot \theta/2 \frac{\partial}{\partial \theta} + \frac{1}{4 \sin^2 \theta/2} \left(\frac{\partial^2}{\partial \nu^2} + \cot \nu \frac{\partial}{\partial \nu} + \frac{1}{\sin^2 \nu} \frac{\partial^2}{\partial \lambda^2} \right) \quad (77)$$

It can be shown that the isotropic solution does not depend on ν or λ , and so all that needs to be solved is [99,100]:

$$\frac{\partial f}{\partial t} = K \left(\frac{\partial^2 f}{\partial \theta^2} + \cot \theta/2 \frac{\partial f}{\partial \theta} \right) \quad (78)$$

subject to initial condition $f(R, 0) = \delta(R)$.

A basis for all functions on $SO(3)$ that depend only on θ are the functions $\{\chi_l(\theta) \mid l \in \mathbb{Z}_{\geq 0}\}$ where

$$\chi_l(\theta) = \frac{\sin(l + \frac{1}{2})\theta}{\sin \frac{\theta}{2}}.$$

These are eigenfunctions of the Laplacian:

$$\nabla^2 \chi = -l(l+1)\chi.$$

Consequently, the Fourier series solution of the isotropic heat equation on $SO(3)$ is known to be:

$$f(R(\theta, \nu, \lambda), t) = \sum_{l=0}^{\infty} (2l+1) \chi_l(\theta) e^{-l(l+1)Kt} = \left(\sin \frac{\theta}{2} \right)^{-1} \sum_{l=0}^{\infty} (2l+1) \sin((l+1/2)\theta) e^{-l(l+1)Kt}. \quad (79)$$

Note that:

$$\lim_{t \rightarrow \infty} f(R(\theta, \nu, \lambda), t) = 1$$

and

$$\int_{SO(3)} f(R, t) dR = 1.$$

When $t = 0$, the above becomes the Fourier series for the Dirac delta function

$$f(R, 0) = \delta(R).$$

As with the case of the heat equation on the circle, an alternative solution exists, analogous to a folded Gaussian. Denote this solution as $\rho(\theta, t)$, and

$$f(R(\theta, \nu, \lambda), t) = e^{Kt/4} \left(\sin \frac{\theta}{2} \right)^{-1} \sum_{k \in \mathbb{Z}} \rho(\theta + 2\pi k, t). \quad (80)$$

This can be derived in two steps. First, let

$$f(R(\theta, \nu, \lambda), t) = \left(\sin \frac{\theta}{2} \right)^{-1} h(\theta, t).$$

Substituting in to (78) and simplifying then gives

$$\frac{\partial h}{\partial t} = K \left(\frac{\partial^2 h}{\partial \theta^2} + \frac{1}{4} h \right). \quad (81)$$

Then substituting $h(\theta, t) = e^{at} q(\theta, t)$ and simplifying gives that when $a = K/4$

$$\frac{\partial q}{\partial t} = K \frac{\partial^2 q}{\partial \theta^2}. \quad (82)$$

The fundamental solution of this heat Equation (the 1D heat kernel) is

$$q(\theta, t) = \frac{1}{2\sqrt{\pi Kt}} e^{-\theta^2/4Kt}.$$

However, this solution is not a satisfactory choice for $\rho(\theta, t)$ for several reasons. First, $SO(3)$ is a 3D space, and so the normalization is not correct, as the normalization factor should be proportional to $1/t^{3/2}$. However, to arbitrarily change the temporal dependence will cause the result to no longer be a solution of (82). Secondly, whereas division by $\sin(\theta/2)$ is fine in the definition of $\chi_l(\theta)$ because zeros are balanced by zeros in the numerator, that is not the case here.

There is a way to solve both problems. That is by realizing that if $q(\theta, t)$ solves (82), then so too does $C\partial q/\partial\theta$ for arbitrary constant C . Consequently, we take as a candidate solution [99,100]:

$$\rho(\theta, t) = C \frac{1}{(\pi Kt)^{3/2}} \theta e^{-\theta^2/4Kt}. \quad (83)$$

The choice of normalizing factor C is made such that (80) is a pdf. This is a constant, hence independent of t . Choosing a relatively small value of t , the summation in (80) reduces to a single term and (76) becomes

$$\frac{1}{\pi} C \frac{e^{Kt/4}}{(\pi Kt)^{3/2}} \int_{-\pi}^{\pi} \theta e^{-\theta^2/4Kt} \sin(\theta/2) d\theta = 1.$$

Moreover, for small t the integral over $[-\pi, \pi]$ can be replaced with an integral over the whole real line. Consequently, since

$$\int_{-\infty}^{\infty} \theta e^{-\theta^2/4Kt} \sin(\theta/2) d\theta = 2\sqrt{\pi}(Kt)^{3/2} e^{-Kt/4}$$

then

$$C = \pi^2/2$$

and

$$f(R(\theta, v, \lambda), t) = \frac{\sqrt{\pi}}{2} \frac{e^{Kt/4}}{(Kt)^{3/2}} \left(\sin \frac{\theta}{2} \right)^{-1} \sum_{k \in \mathbb{Z}} (\theta + 2\pi k) e^{-(\theta + 2\pi k)^2/4Kt}. \quad (84)$$

Since $f(R, t) > 0$ and integrates to 1 it is a valid pdf. What needs to be tested is

$$\lim_{t \rightarrow \infty} f(R, t) = 1$$

and

$$\lim_{t \rightarrow 0} f(R, t) = \delta(R).$$

If these conditions hold, then the solution is valid.

This provides a way to bound entropy in a way similar to the case of the circle. The next section makes more general exact statements for all values of time.

5.4. Rate of Entropy Production under Diffusion on Unimodular Lie Groups

The entropy of a pdf on a Lie group is defined in (5). If $f(g, t)$ is a pdf that satisfies a diffusion equation, then some interesting properties of $S_f(t)$ that are independent of initial conditions result. For example, if $\dot{S}_f = dS_f/dt$, then differentiating under the integral gives

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

Moreover, since f is a pdf,

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

and so the second term in braces in the expression for \dot{S}_f integrates to zero.

Substituting the diffusion equation

$$\frac{\partial f}{\partial t} = \frac{1}{2} \sum_{i,j=1}^n D_{ij} \tilde{E}_i \tilde{E}_j f - \sum_{k=1}^n a_k \tilde{E}_k f \quad (85)$$

into the expression for \dot{S}_f gives [5,6,31]

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

where $F = [F_{ij}]$ is the Lie-group version of the Fisher information matrix with entries

$$F_{ij} \doteq \int_G \frac{1}{f} (\tilde{E}_j f) (\tilde{E}_i f) \, dg.$$

Consequently,

$$\frac{1}{2} \lambda_{\min}(D) \text{tr}[F] \leq \dot{S} \leq \frac{1}{2} \lambda_{\max}(D) \text{tr}[F].$$

The above result is an extension of one presented in [5,6,31].

5.5. The Generalized de Buijn Identity

Here a theorem derived in [5,6,31] is restated.

Theorem 3. Let the solution of the diffusion Equation (85) with constant $\mathbf{a} = [a_1, \dots, a_n]^T$ subject to the initial condition $f(g, 0; D, \mathbf{a}) = \delta(g)$ be denoted as $f_{D,\mathbf{a},t}(g) = f(g, t; D, \mathbf{a})$. Let $\alpha(g)$ be another differentiable pdf on the group. Then

$$\frac{d}{dt} S(\alpha * f_{D,\mathbf{a},t}) = \frac{1}{2} \text{tr}[DF(\alpha * f_{D,\mathbf{a},t})]. \quad (86)$$

Proof. The solution of the diffusion equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{2} \sum_{i,j=1}^n D_{ij} \tilde{E}_i \tilde{E}_j \rho - \sum_{k=1}^n a_k \tilde{E}_k \rho \quad (87)$$

subject to the initial conditions $\rho(g, 0) = \alpha(g)$ is $\rho(g, t) = (\alpha * f_{D,\mathbf{a},t})(g)$. Then computing the derivative of $S(\rho(g, t))$ with respect to time yields

$$\frac{d}{dt} S(\rho) = - \frac{d}{dt} \int_G \rho(g, t) \log \rho(g, t) \, dg = - \int_G \left\{ \frac{\partial \rho}{\partial t} \log \rho + \frac{\partial \rho}{\partial t} \right\} dg. \quad (88)$$

By substituting in (87), the partial with respect to time can be replaced with Lie derivatives. However,

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

Consequently, the second term on the right side of (88) completely disappears. Using the integration-by-parts formula (There are no surface terms. As with the circle and real line, each coordinate in the integral either wraps around or goes to infinity.)

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

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

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

This means that:

$$S(\alpha * f_{D,\mathbf{a},t_2}) - S(\alpha * f_{D,\mathbf{a},t_1}) = \frac{1}{2} \int_{t_1}^{t_2} \text{tr}[D F(\alpha * f_{D,\mathbf{a},t})] dt.$$

□

Whereas some inequalities of Information Theory generalize to the Lie group setting, as demonstrated above, others do not. For example, under convolution on a Lie group (24) and (25) do not hold in general. As for the CRB, there are versions for Lie groups applicable to small values of Dt or with different concepts of covariance, but not in a way that is directly applicable to the scenarios formulated here.

6. Conclusions

Stochastic mechanical systems can describe individual representatives of a statistical mechanical ensemble (as in a rotor in rotational Brownian motion), or can be a stand-alone system subjected to noise, such as a kinematic cart robot. When these systems have mass, Itô and Stratonovich SDEs lead to the same Fokker–Planck equation. Even in the case of inertia-free kinematic systems evolving on Lie groups, these can be the same. From the Fokker–Planck equation, the rate of entropy can be computed. For diffusion processes (with or without drift), the rate of entropy increase is related simply to the diffusion matrix and the Fisher information matrix. This result holds both in Euclidean spaces and on unimodular Lie groups (including cotangent bundle groups), which are a common configuration space for mechanical systems. As systems approach equilibrium, the entropy rate approaches zero. Two different ways to approach equilibrium are discussed: (1) when there is a restoring potential; (2) when the configuration space is bounded. By using the monotonicity and convexity properties of the logarithm function together with inequalities from information theory, computable bounds on entropy and entropy rate are established.

Funding: This work was performed under MOE Tier 1 grant R-265-000-655-114, Faculty Board funds C-265-000-071-001, and National University of Singapore Startup grants R-265-000-665-133 and R-265-000-665-731.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tribus, M. Information theory and thermodynamics. In *Heat Transfer, Thermodynamics and Education: Boelter Anniversary Volume*; Johnson, H.A., Ed.; McGraw-Hill: New York, NY, USA, 1964.
2. Avery, J. *Information Theory and Evolution*; World Scientific: Singapore, 2003.
3. Chirikjian, G.S.; Kyatkin, A.B. *Harmonic Analysis for Engineers and Applied Scientists*; Dover: Mineola, NY, USA, 2016.
4. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University Illinois Press: Urbana, IL, USA, 1949.
5. Chirikjian, G.S. *Stochastic Models, Information Theory, and Lie Groups: Volume I*; Birkhäuser: Boston, MA, USA, 2009.
6. Chirikjian, G.S. *Stochastic Models, Information Theory, and Lie Groups: Volume II*; Birkhäuser: Boston, MA, USA, 2012.
7. Stratonovich, R.L. *Topics in the Theory of Random Noise: Volumes I*; Gordon and Breach Science Publishers, Inc.: New York, NY, USA, 1963.
8. Stratonovich, R.L. *Topics in the Theory of Random Noise: II*; CRC Press: Boca Raton, FL, USA, 1967.
9. Fisher, R. Theory of Statistical Estimation. *Proc. Camb. Philos. Soc.* **1925**, *22*, 700–725. [\[CrossRef\]](#)
10. Kullback, S. *Information Theory and Statistics*; Dover Publications Inc.: Mineola, NY, USA, 1968.
11. Jaynes, E.T. Information Theory and Statistical Mechanics, I. *Phys. Rev.* **1957**, *106*, 620–630. [\[CrossRef\]](#)
12. Jaynes, E.T. Information Theory and Statistical Mechanics, II. *Phys. Rev.* **1957**, *108*, 171–190. [\[CrossRef\]](#)
13. Frieden, B.R. *Physics from Fisher Information*; Cambridge University Press: Cambridge, UK, 1998.
14. Zegers, P. Fisher Information Properties. *Entropy* **2015**, *17*, 4918–4939. [\[CrossRef\]](#)
15. Stam, A.J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control* **1959**, *2*, 101–112. [\[CrossRef\]](#)
16. Blachman, N.M. The convolution inequality for entropy powers. *IEEE Trans. Inform. Theory* **1965**, *11*, 267–271. [\[CrossRef\]](#)
17. Dembo, A.; Cover, T.M.; Thomas, J.A. Information Theoretic Inequalities. *IEEE Trans. Inf. Theory* **1991**, *37*, 1501–1518. [\[CrossRef\]](#)
18. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley-Interscience: Hoboken, NJ, USA, 2006.
19. Muller, S.; Bachmann, M.; Kroninger, D.; Kurz, T.; Helluy, P. Comparison and validation of compressible flow simulations of laser-induced cavitation bubbles. *Comput. Fluids* **2009**, *38*, 1850–1862. [\[CrossRef\]](#)
20. Iwase, M.; Liang, Y.; Masuda, Y.; Morimoto, M.; Matsuoka, T.; Boek, E.S.; Kaito, Y.; Nakagawa, K. Application of a digital oil model to solvent-based enhanced oil recovery of heavy crude oil. *Energy Fuels* **2019**, *33*, 10868–10877. [\[CrossRef\]](#)
21. Jia, X.; Zeng, F.; Gu, Y. Semi-analytical solutions to one-dimensional advection-diffusion equations with variable diffusion coefficient and variable flow velocity. *Appl. Math. Comput.* **2013**, *221*, 268–281. [\[CrossRef\]](#)
22. Orlandi, P.; Bernardini, M.; Pirozzoli, S. Poiseuille and Couette flows in the transitional and fully turbulent regime. *J. Fluid Mech.* **2015**, *770*, 424–441. [\[CrossRef\]](#)
23. Sun, Y.; Jayaraman, A.S.; Chirikjian, G.S. Lie group solutions of advection-diffusion equations. *Phys. Fluids* **2021**, *33*, 046604. [\[CrossRef\]](#)
24. Cramér, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, USA, 1946.
25. Rao, C.R. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–89.
26. Chirikjian, G.S. Information theory on Lie groups and mobile robotics applications. In Proceedings of the 2010 IEEE International Conference on Robotics and Automation, Anchorage, AK, USA, 3–7 May 2010; pp. 2751–2757.
27. Bonnabel, S.; Barrau, A. An intrinsic Cramér-Rao bound on SO(3) for (dynamic) attitude filtering. In Proceedings of the 54th IEEE Conference on Decision and Control (CDC), Osaka, Japan, 15–18 December 2015; pp. 2158–2163.
28. Bonnabel, S.; Barrau, A. An intrinsic Cramér-Rao bound on Lie groups. In Proceedings of the International Conference on Geometric Science of Information, Palaiseau, France, 28–30 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 664–672.
29. Solo, V.; Chirikjian, G.S. On the Cramer-Rao Bound in Riemannian Manifolds with Application to SO(3). In Proceedings of the 2020 59th IEEE Conference on Decision and Control (CDC), Jeju, Korea, 14–18 December 2020; pp. 4117–4122.
30. Solo, V.; Chirikjian, G.S. Ito, Stratonovich and Geometry. In Proceedings of the 2019 IEEE 58th Conference on Decision and Control (CDC), Nice, France, 11–13 December 2019; pp. 3026–3032.
31. Chirikjian, G.S. Information-Theoretic Matrix Inequalities and Diffusion Processes on Unimodular Lie Groups. In *Geometric Structures of Information*; Nielsen, F., Ed.; Springer: Berlin/Heidelberg, Germany, 2019; pp. 213–249.
32. Chirikjian, G.S. From Wirtinger to Fisher Information Inequalities on Spheres and Rotation Groups. In Proceedings of the 21st International Conference on Information Fusion (FUSION), Cambridge, UK, 10–13 July 2018; pp. 730–736.
33. Crassidis, J.L.; Junkins, J.L. *Optimal Estimation of Dynamic Systems*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2004.
34. Gibbs, J.W. *Elementary Principles in Statistical Mechanics: Developed With Especial Reference to the Rational Foundation of Thermodynamics*; Cambridge University Press: Cambridge, UK, 1902; Reissued by Kessinger Publishing and BiblioBazaar 2008.
35. Balescu, R. *Equilibrium and Nonequilibrium Statistical Mechanics*; Wiley: New York, NY, USA, 1975.
36. de Groot, S.R.; Mazur, P. *Non-Equilibrium Thermodynamics*; North-Holland Publishing Co.: Amsterdam, The Netherlands; Interscience Publishers, Inc.: New York, NY, USA, 1962.
37. Prigogine, I. *Non-Equilibrium Statistical Mechanics*; John Wiley and Sons Inc.: New York, NY, USA, 1962.
38. McLennan, J.A. *Introduction to Non-Equilibrium Statistical Mechanics*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, USA, 1989.
39. Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University Press: Oxford, UK, 2001.

40. Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693. [\[CrossRef\]](#)
41. Kac, M. *Some Stochastic Problems in Physics and Mathematics*; Colloquium Lectures in the Pure and Applied Sciences, Magnolia Petroleum Company: Galveston, TX, USA, 1957.
42. Bismut, J.-M. *Mécanique Aléatoire*; Springer: Berlin, Germany, 1981.
43. Nelson, E. *Dynamical Theories of Brownian Motion*; Princeton University Press: Princeton, NJ, USA, 1967.
44. Nelson, E. Review of stochastic mechanics. *J. Physics Conf. Ser.* **2012**, *361*, 012011. [\[CrossRef\]](#)
45. Jayaraman, A.S.; Campolo, D.; Chirikjian, G.S. Black-Scholes theory and diffusion processes on the cotangent bundle of the affine group. *Entropy* **2020**, *22*, 455. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Barbaresco, F. Koszul information geometry and Souriau geometric temperature/capacity of Lie group thermodynamics. *Entropy* **2014**, *16*, 4521–4565. [\[CrossRef\]](#)
47. Barbaresco, F. Geometric theory of heat from Souriau Lie groups thermodynamics and Koszul Hessian geometry: Applications in information geometry for exponential families. *Entropy* **2016**, *18*, 386. [\[CrossRef\]](#)
48. Marle, C.M. From tools in symplectic and Poisson geometry to J.-M. Souriau's theories of statistical mechanics and thermodynamics. *Entropy* **2016**, *18*, 370. [\[CrossRef\]](#)
49. De Saxcé, G. Link between Lie group statistical mechanics and thermodynamics of continua. *Entropy* **2016**, *18*, 254. [\[CrossRef\]](#)
50. Brockett, R.W. System Theory on Group Manifolds and Coset Spaces. *SIAM J. Control* **1972**, *10*, 265–284. [\[CrossRef\]](#)
51. Murray, R.; Li, Z.; Sastry, S. *A Mathematical Introduction to Robotics*; CRC Press: Boca Raton, FL, USA, 1994.
52. Bullo, F.; Lewis, A.D. *Geometric Control of Mechanical Systems*; Springer: Berlin/Heidelberg, Germany, 2004.
53. Holm, D. *Geometric Mechanics, Part I: Dynamics and Symmetry*; World Scientific: Singapore, 2008.
54. Holm, D. *Geometric Mechanics, Part II: Rotating, Translating and Rolling*; World Scientific: Singapore, 2008.
55. Smith, P.; Drummond, T.; Roussopoulos, K. Computing MAP Trajectories by Representing, Propagating and Combining PDFs over Groups. In Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; Volume 2, pp. 1275–1282.
56. Thrun, S.; Burgard, W.; Fox, D. *Probabilistic Robotics*; MIT Press: Cambridge, MA, USA, 2005.
57. Park, W.; Liu, Y.; Moses, M.; Chirikjian, G.S. Kinematic State Estimation and Motion Planning for Stochastic Nonholonomic Systems Using the Exponential Map. *Robotica* **2008**, *26*, 419–434. [\[CrossRef\]](#)
58. Zhou, Y.; Chirikjian, G.S. Probabilistic Models of Dead-Reckoning Error in Nonholonomic Mobile Robots. In Proceedings of the 2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422), Taipei, Taiwan, 14–19 September 2003.
59. Long, A.W.; Wolfe, K.C.; Mashner, M.J.; Chirikjian, G.S. The banana distribution is Gaussian: A localization study with exponential coordinates. In *Robotics: Science and Systems VIII*; MIT Press: Cambridge, MA, USA, 2013; p. 265.
60. Mumford, D. Elastica and computer vision. In *Algebraic Geometry and Its Applications*; Bajaj, C., Ed.; Springer: New York, NY, USA, 1994.
61. Williams, L.R.; Jacobs, D.W. Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Comput.* **1997**, *9*, 837–858. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Williams, L.R.; Jacobs, D.W. Local Parallel Computation of Stochastic Completion Fields. *Neural Comput.* **1997**, *9*, 859–881. [\[CrossRef\]](#)
63. Zweck, J.; Williams, L.R. Euclidean Group Invariant Computation of Stochastic Completion Fields Using Shiftable-Twistable Functions. *J. Math. Imaging Vis.* **2004**, *21*, 135–154. [\[CrossRef\]](#)
64. Citti, G.; Sarti, A. A Cortical Based Model of Perceptual Completion in the Roto-Translation Space. *J. Math. Imaging Vis.* **2006**, *24*, 307–326. [\[CrossRef\]](#)
65. Duits, R.; Franken, E. Left-invariant parabolic evolutions on $SE(2)$ and contour enhancement via invertible orientation scores Part I: Linear left-invariant diffusion equations on $SE(2)$. *Quart. Appl. Math.* **2010**, *68*, 255–292. [\[CrossRef\]](#)
66. Bond, D.J. The statistical properties of phase noise. *Br. Telecom. Technol. J.* **1989**, *7*, 12–17.
67. Foschini, G.J.; Greenstein, L.J.; Vannucci, G. Noncoherent detection of coherent lightwave signals corrupted by phase noise. *IEEE Trans. Commun.* **1988**, COM-36, 306–314. [\[CrossRef\]](#)
68. Foschini, G.J.; Vannucci, G. Characterizing filtered light waves corrupted by phase noise. *IEEE Trans. Inf. Theory* **1988**, *34*, 1437–1448. [\[CrossRef\]](#)
69. Foschini, G.J.; Vannucci, G.; Greenstein, L.J. Envelope statistics for filtered optical signals corrupted by phase noise. *IEEE Trans. Commun.* **1989**, *37*, 1293–1302. [\[CrossRef\]](#)
70. Garrett, I.; Bond, D.J.; Waite, J.B.; Lettis, D.S.L.; Jacobsen, G. Impact of phase noise in weakly coherent systems: A new and accurate approach. *J. Light. Technol.* **1990**, *8*, 329–337. [\[CrossRef\]](#)
71. Garrett, I.; Jacobsen, G. Phase noise in weakly coherent systems. *IEEE Proc.* **1989**, *136 Pt. J*, 159–165. [\[CrossRef\]](#)
72. Zhang, X. Analytically solving the Fokker–Planck equation for the statistical characterization of the phase noise in envelope detection. *J. Light. Technol.* **1995**, *13*, 1787–1794. [\[CrossRef\]](#)
73. Wang, Y.; Zhou, Y.; Maslen, D.K.; Chirikjian, G.S. Solving the Phase-Noise Fokker–Planck Equation Using the Motion-Group Fourier Transform. *IEEE Trans. Commun.* **2006**, *54*, 868–877. [\[CrossRef\]](#)
74. 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. In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 4600–4605.

75. Park, W.; Wang, Y.; Chirikjian, G.S. The path-of-probability algorithm for steering and feedback control of flexible needles. *Int. J. Robot. Res.* **2010**, *29*, 813–830. [[CrossRef](#)] [[PubMed](#)]
76. Webster, R.J., III; Kim, J.-S.; Cowan, N.J.; Chirikjian, G.S.; Okamura, A.M. Nonholonomic Modeling of Needle Steering. *Int. J. Robot. Res.* **2006**, *25*, 509–525. [[CrossRef](#)]
77. Perrin, F. “Étude Mathématique du Mouvement Brownien de Rotation. *Ann. Sci. L'École Norm. Supérieure* **1928**, *45*, 1–51. [[CrossRef](#)]
78. Furry, W.H. Isotropic Rotational Brownian Motion. *Phys. Rev.* **1957**, *107*, 7–13. [[CrossRef](#)]
79. Favro, L.D. Theory of the Rotational Brownian Motion of a Free Rigid Body. *Phys. Rev.* **1960**, *119*, 53–62. [[CrossRef](#)]
80. Hubbard, P.S. Angular velocity of a nonspherical body undergoing rotational Brownian motion. *Phys. Rev. A* **1977**, *15*, 329–336. [[CrossRef](#)]
81. Steele, W.A. Molecular Reorientation in Liquids. I. Distribution Functions and Friction Constants. II. Angular Autocorrelation Functions. *J. Chem. Phys.* **1963**, *38*, 2404–2418. [[CrossRef](#)]
82. McConnell, J. *Rotational Brownian Motion and Dielectric Theory*; Academic Press: New York, NY, USA, 1980.
83. Weber, G. Rotational Brownian motion and polarization of the fluorescence of solutions. *Adv. Protein Chem.* **1953**, *8*, 415–459. [[PubMed](#)]
84. Tao, T. Time-dependent fluorescence depolarization and Brownian rotational diffusion coefficients of macromolecules. *Biopolymers* **1969**, *8*, 609–632. [[CrossRef](#)]
85. Willsky, A.S. Dynamical Systems Defined on Groups: Structural Properties and Estimation. Ph.D. Dissertation, Department Aeronautics and Astronautics, M.I.T., Cambridge, MA, USA, June 1973.
86. Itô, K. Brownian Motions in a Lie Group. *Proc. Jpn. Acad.* **1950**, *26*, 4–10. [[CrossRef](#)]
87. Itô, K. Stochastic Differential Equations in a Differentiable Manifold. *Nagoya Math. J.* **1950**, *1*, 35–47. [[CrossRef](#)]
88. Itô, K. Stochastic Differential Equations in a Differentiable Manifold (2). *Mem. Coll. Sci. Univ. Kyoto Ser. A Math.* **1953**, *28*, 81–85. [[CrossRef](#)]
89. McKean, H.P., Jr. *Stochastic Integrals*; Academic Press: New York, NY, USA, 1969.
90. Gangolli, R. On the Construction of Certain Diffusions on a Differentiable Manifold. *Z. Wahrscheinlichkeitstheorie Und Verw. Geb.* **1964**, *2*, 406–419. [[CrossRef](#)]
91. Duncan, T.E. Stochastic Systems in Riemannian Manifolds. *J. Optim. Theory Appl.* **1979**, *27*, 399–426. [[CrossRef](#)]
92. McKean, H.P., Jr. Brownian Motions on the 3-Dimensional Rotation Group. *Mem. Coll. Sci. Univ. Kyoto Ser. A* **1960**, *33*, 25–38. [[CrossRef](#)]
93. Gorman, C.D. Brownian Motion of Rotation. *Trans. Am. Soc.* **1960**, *94*, 103–117. [[CrossRef](#)]
94. Liao, M. Random Motion of a Rigid Body. *J. Theor.* **1997**, *10*, 201–211.
95. Liao, M. *Lévy Processes in Lie Groups*; Cambridge University Press: Cambridge, UK, 2004.
96. Jeffrey, G.B. The Motion of Ellipsoidal Particles Immersed in a Viscous Fluid. *Proc. R. Soc. Lond. Ser. A* **1922**, *102*, 161–179.
97. Varshalovich, D.A.; Moskalev, A.N.; Khersonskii, V.K. *Quantum Theory of Angular Momentum*; World Scientific: Singapore, 1988.
98. Gel'f, I.M.; Minlos, R.A.; Shapiro, Z.Y. *Representations of the Rotation and Lorentz Groups and Their Applications*; Macmillan: New York, NY, USA, 1963.
99. Chétalet, O.; Chirikjian, G.S. Sampling and Convolution on Motion Groups Using Generalized Gaussian Functions. *Electron. Comput. Kinemat.* **2002**, *1*.
100. Lee, S.; Chirikjian, G.S. Pose Analysis of Alpha-Carbons in Proteins. *Int. J. Robot. Res.* **2005**, *24*, 183–210. [[CrossRef](#)]