Peer-reviewed version available at Entropy 2018, 20, 696; doi:10.3390/e20090696

Probabilistic inference for dynamical systems

Sergio Davis,[1,*] Diego González,[1,2] and Gonzalo Gutiérrez[2]

[1] *Comisión Chilena de Energía Nuclear, Casilla 188-D, Santiago, Chile*

[2] *Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile*

Abstract

A general framework for inference in dynamical systems is described, based on the language of Bayesian probability theory and making use of the maximum entropy principle. Taking as fundamental the concept of a path, the continuity equation and Cauchy's equation for fluid dynamics arise naturally, while the specific information about the system can be included using the Maximum Caliber (or maximum path entropy) principle.

* sergio.davis@cchen.cl

I. INTRODUCTION

Dynamical system models are widely used to describe complex physical systems (such as the weather), and also social and economic systems (such as the stock market). These systems are usually subject to high levels of uncertainty, either in their initial conditions and/or in their interactions with their own environment. From the point of view of constructing predictive models, the optimal description of the time-dependent state of such a system given external constraints is a challenge with promising applications in both fundamental and applied Science. This is of course an *inference problem* in which we must choose the most likely solution out of the (possibly infinite) alternatives compatible with the given information we have about the system.

Given all this, it seems that a unified framework for performing inference on dynamical systems may open new possibilities in several areas, including non-equilibrium statistical mechanics and thermodynamics, hydrodynamics (including magnetohydrodynamics) and classical mechanics under stochastic forces, among other possible fields of application.

In this work, we present some elements for a general framework of inference in dynamical systems, written in the language of Bayesian probability. The first is the master equation, which is shown as a direct consequence of the laws of probability. Next we develop the treatment of inference over paths from which we obtain the continuity equation and Cauchy's equation for fluid dynamics, and discuss their range of applicability. Finally we close with some concluding remarks.

II. WHY BAYESIAN INFERENCE?

Unlike the standard ("frequentist") interpretation of probability theory, in which probabilities are frequencies of occurrence of repeatable events, Bayesian probability can be understood as the natural extension of classical logic in the case of *uncertainty* [1, 2]. Bayesian probability deals with unknown quantities rather than identical repetitions/copies of an event or system, and is able to include *prior* information when needed.

The conceptual framework of Bayesian probability provides an elegant language to describe dynamical systems under uncertainty. A straightforward advantage of the Bayesian framework is that one does not need to assume an ensemble of "many identical copies of

a system", a single system with uncertain initial conditions and/or forces is sufficient to construct a theory. The probability of finding this particular system in a given state at a given time would not be a frequency, but rather a degree of plausibility conditioned on the known information. In fact, we can lift even the common assumption of "many degrees of freedom", the motion of a single particle could be used to construct an internally consistent theory, where the equations of motion for time-dependent probability densities are similar to the ones of hydrodynamics. We will describe in detail both features of the Bayesian formulation in the following sections.

A brief overview of Bayesian notation used in this work follows. We will take $P(\mathcal{Q}|\mathcal{I})$ as the probability of a particular *proposition* $\mathcal{Q}$ being true given knowledge $\mathcal{I}$. On the other hand, $\left\langle G\right\rangle_{\mathcal{I}}$ will denote the expected value of an arbitrary *quantity* G given knowledge $\mathcal{I}$, and will be given by

$$\left\langle G\right\rangle_{\mathcal{I}} = \sum_{\boldsymbol{u}} P(\boldsymbol{u}|\mathcal{I})G(\boldsymbol{u}), \tag{1}$$

where $\boldsymbol{u}$ represents one of the possible states of the system.

III. UNIVERSALITY OF THE MASTER EQUATION

Consider a discrete-time system which can transit between n possible states $\{x_1, \ldots, x_n\}$ at different times. If we denote by $X(t)$ the state of the system at time t, we have that the joint probability of being in state a at time t and in state b at time $t + \Delta t$ is given by

$$P(X(t) = a, X(t + \Delta t) = b|\mathcal{I}) = P(X(t + \Delta t) = b|X(t) = a, \mathcal{I}) \cdot P(X(t) = a|\mathcal{I}). \tag{2}$$

By summing over a in Eq. 2 and taking $b = x$ we have

$$P(X(t + \Delta t) = x|\mathcal{I}) = \sum_{a} P(X(t + \Delta t) = x|X(t) = a, \mathcal{I}) \cdot P(X(t) = a|\mathcal{I}), \tag{3}$$

while, by summing over b and taking $a = x$ we have, on the other hand

$$P(X(t) = x|\mathcal{I}) = P(X(t) = x|\mathcal{I}) \sum_{b} P(X(t + \Delta t) = b|X(t) = x, \mathcal{I}). \tag{4}$$

From these two identities (Eqs. 3 and 4) we can construct the discrete-time difference,

$$\frac{P(X(t+\Delta t)=x|\mathcal{I})-P(X(t)=x|\mathcal{I})}{\Delta t}=\sum_{x'}\left[\frac{P(X(t+\Delta t)=x|X(t)=x',\mathcal{I})}{\Delta t}\right]P(X(t)=x'|\mathcal{I}) - P(X(t)=x|\mathcal{I})\sum_{x'}\left[\frac{P(X(t+\Delta t)=x'|X(t)=x,\mathcal{I})}{\Delta t}\right], \tag{5}$$

which is the celebrated *master equation* [3–5]. In the continuous-time limit, i.e. when $\Delta t \to 0$, we can write this master equation as

$$\frac{\partial}{\partial t}\rho(x;t)=\sum_{x'}W_t(x'\to x)\rho(x';t)-\rho(x;t)\sum_{x'}W_t(x\to x'). \tag{6}$$

where we have defined the (continuous-time) *transition rate* W as

$$W_t(a\to b):=\lim_{\Delta t\to 0}\frac{P(X(t+\Delta t)=b|X(t)=a,\mathcal{I})}{\Delta t} \tag{7}$$

and the instantaneous density ρ as $\rho(a;t):=P(X(t)=a|\mathcal{I})$. In this sense, the master equation is a direct consequence of the laws of probability, its validity is universal whenever we have transitions between states. It follows from this that all probabilities of time-dependent quantities must evolve in time according to Eq. 5 or (6 in the case of continuous time) for some transition probability (rate). The master equation is more general than the continuity equation, as it includes the case where some quantities can be created or destroyed during a process. However, time evolution under global and local conservation laws is a fundamental case that can also be readily obtained from the Bayesian formalism, as we will see in the following sections. As it is well known, the continuous-time master equation can be approximated in the limit of infinitesimally small transitions to obtain the Fokker-Planck equation [4, 6].

IV. FLUID THEORIES IN A BAYESIAN FORMULATION

Now we will consider a dynamical system that follows a path $\boldsymbol{X}(t) \in \mathbb{X}$ in time, where $\mathbb{X}$ denotes the space of all paths consistent with given boundary conditions. The path $\boldsymbol{X}(t)$ is not completely known, and we have only access to partial information denoted by $\mathcal{I}$.

Peer-reviewed version available at Entropy 2018, 20, 696; doi:10.3390/e20090696

In this setting, Bayesian theory defines a functional $P[\boldsymbol{X}|\mathcal{I}]$ which is the probability density of the path $\boldsymbol{X}(t)$ being the "true path" under the known information. For any arbitrary *functional* $\mathcal{F}[\boldsymbol{X}]$ of the path we can then write its expected value as a path integral

$$\Big\langle \mathcal{F} \Big\rangle_{\mathcal{I}} = \int_{\mathbb{X}} \mathcal{D}\boldsymbol{X} P[\boldsymbol{X}|\mathcal{I}] \mathcal{F}[\boldsymbol{X}]. \tag{8}$$

On the other hand, the expected value of any instantaneous *quantity* $A(\boldsymbol{x};t)$ is given by

$$\Big\langle A \Big\rangle_{t,\mathcal{I}} = \int_{\mathbb{X}} \mathcal{D}\boldsymbol{X} P[\boldsymbol{X}|\mathcal{I}] A(\boldsymbol{X}(t);t) = \int_{V} d\boldsymbol{x} \rho(\boldsymbol{x};t) A(\boldsymbol{x};t), \tag{9}$$

where $\rho(\boldsymbol{x};t) := P(\boldsymbol{X}(t) = \boldsymbol{x}|\mathcal{I})$ is the instantaneous probability density at time t.

By using a quantity $A(\boldsymbol{x'};t) = \delta(\boldsymbol{x'} - \boldsymbol{x})$ we see that the probability density itself has a path integral representation,

$$\rho(\boldsymbol{x};t) = \int_{\mathbb{X}} \mathcal{D}\boldsymbol{X} P[\boldsymbol{X}|\mathcal{I}] \delta(\boldsymbol{X}(t) - \boldsymbol{x}). \tag{10}$$

By differentiating Eq. 10 with respect to time, we obtain the continuity equation for the instantaneous probability density [7],

$$\begin{aligned}
\frac{\partial \rho}{\partial t} &= \int_{\mathbb{X}} \mathcal{D}\boldsymbol{X} P[\boldsymbol{X}|\mathcal{I}] \frac{\partial}{\partial t} \delta(\boldsymbol{X}(t) - \boldsymbol{x}) \\
&= \int_{\mathbb{X}} \mathcal{D}\boldsymbol{X} P[\boldsymbol{X}|\mathcal{I}] \nabla \delta(\boldsymbol{X}(t) - \boldsymbol{x}) \cdot \dot{\boldsymbol{X}}(t) \\
&= -\nabla \cdot \Big(\int_{\mathbb{X}} \mathcal{D}\boldsymbol{X} P[\boldsymbol{X}|\mathcal{I}] \delta(\boldsymbol{X}(t) - \boldsymbol{x}) \dot{\boldsymbol{X}}(t) \Big) \\
&= -\nabla \cdot \Big(\rho \cdot \boldsymbol{v}(\boldsymbol{x};t) \Big),
\end{aligned} \tag{11}$$

where $\boldsymbol{v}(\boldsymbol{x};t)$ is the velocity field that describes the flow of probability, given by

$$\boldsymbol{v}(\boldsymbol{x};t) = \Big\langle \dot{\boldsymbol{X}}(t) \Big\rangle_{\boldsymbol{X}(t)=\boldsymbol{x},\mathcal{I}} = \frac{1}{\rho(\boldsymbol{x};t)} \Big\langle \dot{\boldsymbol{X}}(t) \delta(\boldsymbol{X}(t) - \boldsymbol{x}) \Big\rangle_{\mathcal{I}}. \tag{12}$$

This equation describes the global and local conservation of the probability of finding the system in a given state $\boldsymbol{x}$ at a time t, and is guaranteed to hold for any system moving continuously in time through paths $\boldsymbol{X}(t) \in \mathbb{X}$.

In the same way, it is possible to derive a dynamical equation for the velocity field $\boldsymbol{v}(\boldsymbol{x};t)$ itself, by differentiating it with respect to time. We have

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 May 2018 doi:10.20944/preprints201805.0028.v1

Peer-reviewed version available at Entropy 2018, 20, 696; doi:10.3390/e20090696

$$\begin{aligned}\frac{\partial}{\partial t}(\rho\cdot v_\mu) &= \frac{\partial}{\partial t}\int_{\mathbb{X}}\mathcal{D}\boldsymbol{X}\,P[\boldsymbol{X}|\mathcal{I}]\delta(\boldsymbol{X}(t)-\boldsymbol{x})\dot{X}_\mu(t)\\ &= \int_{\mathbb{X}}\mathcal{D}\boldsymbol{X}\,P[\boldsymbol{X}|\mathcal{I}]\frac{\partial}{\partial t}\delta(\boldsymbol{X}(t)-\boldsymbol{x})\dot{X}_\mu(t)\\ &= \int_{\mathbb{X}}\mathcal{D}\boldsymbol{X}\,P[\boldsymbol{X}|\mathcal{I}]\Big(\delta(\boldsymbol{X}(t)-\boldsymbol{x})\ddot{X}_\mu(t)+\partial_\nu\delta(\boldsymbol{X}(t)-\boldsymbol{x})\dot{X}_\mu(t)\dot{X}_\nu(t)\Big)\\ &= \Big\langle\delta(\boldsymbol{X}(t)-\boldsymbol{x})\ddot{X}_\mu(t)\Big\rangle_{\mathcal{I}}-\partial_\nu\Big\langle\delta(\boldsymbol{X}(t)-\boldsymbol{x}))\dot{X}_\mu(t)\dot{X}_\nu(t)\Big\rangle_{\mathcal{I}}.\end{aligned}\tag{13}$$

By using the continuity equation (Eq. 11) to replace $\partial\rho/\partial t$, and after some rearrangements, we arrive at the Cauchy momentum equation

$$\frac{\partial\boldsymbol{v}}{\partial t}+\boldsymbol{v}\cdot\nabla\boldsymbol{v}=\frac{D\boldsymbol{v}}{Dt}=\boldsymbol{a}(\boldsymbol{x};t)+\frac{1}{\rho}\nabla\cdot\overleftrightarrow{\boldsymbol{\sigma}}\tag{14}$$

where

$$\boldsymbol{a}(\boldsymbol{x};t):=\Big\langle\ddot{\boldsymbol{X}}(t)\Big\rangle_{\boldsymbol{X}(t)=\boldsymbol{x},\mathcal{I}}\tag{15}$$

is the acceleration field, D/Dt is the *convective derivative* and

$$\overleftrightarrow{\boldsymbol{\sigma}}=-\rho\cdot\overleftrightarrow{\boldsymbol{C}}\tag{16}$$

is the stress tensor. The symmetric tensor $\overleftrightarrow{\boldsymbol{C}}$ is the velocity covariance matrix at a given point, defined by

$$C_{\mu\nu}(\boldsymbol{x};t)=\Big\langle\dot{X}_\mu(t)\dot{X}_\nu(t)\Big\rangle_{\boldsymbol{X}(t)=\boldsymbol{x},\mathcal{I}}-v_\mu(\boldsymbol{x};t)v_\nu(\boldsymbol{x};t).\tag{17}$$

Both Eq. 11 and Eq. 14 form a closed coupled system of equations for $\rho(\boldsymbol{x};t)$ and $\boldsymbol{v}(\boldsymbol{x};t)$, needing as their only external input the velocity covariance matrix $C_{\mu\nu}$. These equations are then *built-in features* of inference over paths. In a Bayesian approach, they are valid for any system that moves continuously in time. The Cauchy momentum equation includes most notably the Navier-Stokes equation as a particular case [8].

V. INCLUDING PARTICULAR KNOWLEDGE INTO OUR MODELS

At this point, we have developed a generic framework where no particular details about a system have been included. Clearly all those details have to be contained in $P[\boldsymbol{X}|\mathcal{I}]$,

Peer-reviewed version available at Entropy 2018, 20, 696; doi:10.3390/e20090696

or rather, in the covariance matrix $C_{\mu\nu}(\boldsymbol{x};t)$ which can be derived from it. The question remains about how to incorporate these details in the most unbiased manner. We could, in principle, start from the *null* assumption of equiprobable paths,

$$P[\boldsymbol{X}|\mathcal{I}_0] = \text{constant},$$

and add new information $\mathcal{R}$ later on, by updating our probability functional $P[\boldsymbol{X}|\mathcal{I}_0]$ to a new $P[\boldsymbol{X}|\mathcal{I}]$, where $\mathcal{I} = (\mathcal{I}_0, \mathcal{R})$. There are essentially two equivalent methods to achieve this, and depending on the actual form of $\mathcal{R}$ one of them may be more directly applicable than the other,

(1) **Bayes' theorem**: the posterior distribution $P(\boldsymbol{u}|\mathcal{I}_0, \mathcal{R})$ is given in terms of the prior $P(\boldsymbol{u}|\mathcal{I}_0)$ by

$$P(\boldsymbol{u}|\mathcal{I}_0, \mathcal{R}) = \frac{P(\boldsymbol{u}|\mathcal{I}_0) \cdot P(\mathcal{R}|\boldsymbol{u}, \mathcal{I}_0)}{P(\mathcal{R}|\mathcal{I}_0)}.$$

This method is most useful when $\mathcal{R}$ is comprised of statements about the states $\boldsymbol{u}$, for instance, boundary conditions.

(2) **Principle of maximum entropy**: The posterior distribution $p(\boldsymbol{u})$ is the one that maximizes

$$\mathcal{S}[p_0 \to p] = -\int_V d\boldsymbol{u}\; p(\boldsymbol{u}) \ln \frac{p(\boldsymbol{u})}{p_0(\boldsymbol{u})}$$

where $p_0(\boldsymbol{u})$ is the prior distribution. This method is most useful when $\mathcal{R}$ consists of constraints on the final model $p(\boldsymbol{u})$, usually expressed as fixed expected values.

VI. THE MAXIMUM CALIBER PRINCIPLE

The function $p(\boldsymbol{u})$ that is closest to our prior probability $p_0(\boldsymbol{u})$ and is consistent with the constraints $\mathcal{R}$ is the one that maximizes the relative entropy [9, 10]

$$\mathcal{S}[p_0 \to p] = -\int_V d\boldsymbol{u} p(\boldsymbol{u}) \ln \frac{p(\boldsymbol{u})}{p_0(\boldsymbol{u})}$$

among the set of functions p that are compatible with $\mathcal{R}$. The negative of this relative entropy measures the "informational distance" from p_0 to p.

It is important to note that this is a rule of inference and not a physical principle, and therefore it is not bounded by the meaning assigned to the states $\boldsymbol{x}$, as long as we can write (Bayesian) probabilities over them.

For the general case of m constraints of the form

$$\Big\langle f_j(\boldsymbol{u}) \Big\rangle = F_j, \qquad (\, j = 1, \ldots, m)$$

the maximum entropy solution starting from $P(\boldsymbol{u}|\mathcal{I}_0)$ is obtained through the use of m Lagrange multipliers (one for each constraint),

$$P(\boldsymbol{u}|\mathcal{I}_0, \mathcal{R}) = P(\boldsymbol{u}|\mathcal{I}_0) \left[\frac{\exp\Big(-\sum_{j=1}^{m} \lambda_j f_j(\boldsymbol{x}) \Big)}{Z(\boldsymbol{\lambda})} \right], \tag{18}$$

where $Z(\boldsymbol{\lambda})$ is the partition function. This is compatible with Bayesian updating, as this posterior distribution is proportional to the prior. The Lagrange multipliers are solutions of the constraint equations in terms of Z,

$$F_j = \frac{\partial}{\partial \lambda_j} \ln Z(\boldsymbol{\lambda}) \qquad (\, j = 1, \ldots, m). \tag{19}$$

In exactly the same way, the path (relative) entropy (sometimes known as the *caliber*) is defined as the path integral [11–18]

$$\mathcal{S}[p_0 \to p] = -\int_{\mathbb{X}} D\boldsymbol{X}\; p[\boldsymbol{X}] \ln \frac{p[\boldsymbol{X}]}{p_0[\boldsymbol{X}]},$$

where $p_0[\boldsymbol{X}] := P[\boldsymbol{X}|\mathcal{I}_0]$ is the prior path probability. The use of this generalization is justified based on the fact that we can write any path $\boldsymbol{X}(t)$ in terms of a complete orthonormal basis $\{\boldsymbol{B}_i\}$,

$$\boldsymbol{X}(t) = \sum_{n=0}^{N-1} \gamma_n \boldsymbol{B}_n(t), \qquad N \to \infty \tag{20}$$

and then there is a one-to-one correspondence between every path $\boldsymbol{X}(t)$ and its coordinates $(\gamma_0, \gamma_1, \ldots, \gamma_{N-1})$. Inference over paths $\boldsymbol{X}$ then becomes completely equivalent to inference over the coefficients $\boldsymbol{\gamma}$, which form a system with N degrees of freedom.

In summary, for the general maximum caliber inference problem we have m constraints, written as

$$\Big\langle \mathcal{F}_j[\boldsymbol{X}] \Big\rangle = F_j \qquad (\, j = 1, \ldots, m), \tag{21}$$

from which the probability functional obtained is

$$P[\boldsymbol{X}|\mathcal{I}_0, \mathcal{R}] = \frac{1}{Z(\boldsymbol{\lambda})} P[\boldsymbol{X}|\mathcal{I}_0] \exp\Big(- \sum_{j=1}^{m} \lambda_j \mathcal{F}_j[\boldsymbol{X}] \Big). \tag{22}$$

Any such maximum caliber solution can be cast in the "canonical" form, as

$$P[\boldsymbol{X}|\mathcal{I}] = \frac{1}{Z(\alpha)} \exp\Big(- \frac{1}{\alpha} \mathcal{A}[\boldsymbol{X}] \Big), \tag{23}$$

where $\mathcal{A}$ is a functional, analogous to the Hamilton action of a classical system, and $\alpha > 0$ is a constant with the same physical units as $\mathcal{A}$. By simple inspection of this canonical form, it is straightforward to see that the most probable path is the one with minimum action

$$\frac{\delta \mathcal{A}[\boldsymbol{X}]}{\delta \boldsymbol{X}(t)} = 0 \qquad \forall t. \tag{24}$$

VII. AN ILLUSTRATION: NEWTONIAN MECHANICS OF CHARGED PARTICLES

As an example of the application of this formalism, consider a "particle" with known square speed ν^2, known instantaneous probability density ρ and known velocity field $\boldsymbol{v}(\boldsymbol{x}; t)$ for all times $t \in [0, \tau]$. The corresponding constraints are then

$$\Big\langle \Big| \dot{\boldsymbol{X}}(t) \Big|^2 \Big\rangle = \nu^2(t) \qquad \forall t, \tag{25}$$

$$\Big\langle \delta(\boldsymbol{X}(t) - \boldsymbol{x}) \Big\rangle = \rho(\boldsymbol{x}; t) \qquad \forall \boldsymbol{x}, t, \tag{26}$$

$$\Big\langle \dot{\boldsymbol{X}}(t) \delta(\boldsymbol{X}(t) - \boldsymbol{x}) \Big\rangle = \rho(\boldsymbol{x}; t) \boldsymbol{v}(\boldsymbol{x}; t) \qquad \forall \boldsymbol{x}, t. \tag{27}$$

The resulting maximum caliber solution is of the form

$$P[\boldsymbol{X}|\mathcal{I}] = \frac{1}{Z(\alpha)} \exp\Big(- \frac{1}{\alpha} \int_0^\tau dt \; L(\boldsymbol{X}(t), \dot{\boldsymbol{X}}(t); t) \Big), \tag{28}$$

with the Hamilton action

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 May 2018 doi:10.20944/preprints201805.0028.v1

Peer-reviewed version available at Entropy 2018, 20, 696; doi:10.3390/e20090696

$$\mathcal{A}[\boldsymbol{X}] = \int_0^\tau dt L(\boldsymbol{X}(t), \dot{\boldsymbol{X}}(t); t), \tag{29}$$

and a Lagrangian defined as

$$L(\boldsymbol{X}(t), \dot{\boldsymbol{X}}(t); t) = \lambda_1(t)|\dot{\boldsymbol{X}}(t)|^2 + \int_V d\boldsymbol{x} \left(\lambda_2(\boldsymbol{x}; t) + \boldsymbol{\lambda_3}(\boldsymbol{x}; t) \cdot \dot{\boldsymbol{X}}(t)\right) \delta(\boldsymbol{X}(t) - \boldsymbol{x}) \tag{30}$$

where λ_1, λ_2 and $\boldsymbol{\lambda_3}$ are Lagrange multipliers. This Lagrangian can be cast into a more familiar form,

$$L(\boldsymbol{X}(t), \dot{\boldsymbol{X}}(t); t) = \frac{1}{2} m(t) \dot{\boldsymbol{X}}(t)^2 - \Phi(\boldsymbol{X}(t); t) + \dot{\boldsymbol{X}}(t) \cdot \boldsymbol{A}(\boldsymbol{X}(t); t) \tag{31}$$

by simply renaming the Lagrange multipliers and integrating the delta function [14]. Interestingly, this is none other than the Lagrangian for a particle with time-dependent "mass" $m(t)$ in an external "electromagnetic potential" $(\Phi, \boldsymbol{A})$. The most probable path under these constraints will of course follow Newton's second law under a "Lorenz force"

$$\frac{d}{dt}\left(m\boldsymbol{v}\right) = -\nabla\Phi(\boldsymbol{x}; t) - \frac{\partial \boldsymbol{A}(\boldsymbol{x}; t)}{\partial t} + \boldsymbol{v} \times (\nabla \times \boldsymbol{A}(\boldsymbol{x}; t)). \tag{32}$$

In particular, it is important to note that it is the constraint on the squared speed $\nu^2(t)$ that adds the mass $m(t)$ to the model, as $m = 2\lambda_1$, the constraint on the probability density $\rho(\boldsymbol{x}; t)$ adds the scalar potential $\Phi(\boldsymbol{x}; t)$ to the model, as $\Phi = -\lambda_2$, and finally the constraint on the local velocity field $\boldsymbol{v}(\boldsymbol{x}; t)$ adds the vector potential $\boldsymbol{A}(\boldsymbol{x}; t)$ to the model, because $\boldsymbol{A} = \boldsymbol{\lambda}_3$. Nowhere in the derivation of this Lagrangian have we assumed the existence of charges, electromagnetic fields, or the Lorenz force, the structure that is revealed is the most unbiased under the constraints given in Eqs. 25 to 27, i.e. with approximate knowledge of its location (given by ρ) and velocity "field lines" (given by $\boldsymbol{v}$)). This model could be used for people in a busy street crossing, or vehicles in a city.

VIII. CONCLUDING REMARKS

We have shown that it is possible to construct a fluid theory from Bayesian inference of an abstract system with N degrees of freedom moving along paths $\boldsymbol{X}(t)$, and that this theory automatically includes the continuity equation and the Cauchy momentum equation

as built-in features. Moreover, through the use of the Maximum Caliber principle it is possible to formulate the dynamics of such an abstract system in terms of an Action that is minimal for the most probable path, resembling the well-known structures of Lagrangian and Hamiltonian mechanics.

By entering into our model the square speed, instantaneous probability density and velocity field, a Lagrangian of a "particle" under external fields emerges naturally. This "particle" moves in average according to Newton's law of motion $\boldsymbol{F} = m \cdot \boldsymbol{a}$ under the Lorenz force, with scalar and vector potentials determined by the known information about the location and velocity lines. In this formulation the only ingredients that we could call *physical* have been the existence of an N-dimensional "particle" moving continuously along (unknown) paths. In this application we see that position and velocity are the only intrinsic (*real* or *ontological*) properties of the particle at a given time t. On the other hand, the time-dependent mass $m(t)$ and the fields $\Phi(\boldsymbol{x};t)$ and $\boldsymbol{A}(\boldsymbol{x};t)$ are emergent parameters (in fact, Lagrange multipliers) needed to impose the constraints on the known information used to construct the model.

IX. ACKNOWLEDGEMENTS

SD and GG gratefully acknowledge funding from FONDECYT 1171127.

[1] R. T. Cox, Am. J. Phys. **14**, 1 (1946).

[2] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).

[3] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, 1996).

[4] N. G. van Campen, *Stochastic processes in Physics and Chemistry* (North Holland, 2007).

[5] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, 2001).

[6] N. G. van Campen, Adv. Chem. Phys **34**, 245 (1976).

[7] D. González, D. Díaz, and S. Davis, Eur. Phys. J. B **89**, 214 (2016).

[8] H. Lamb, *Hydrodynamics*, Dover Books on Physics (Dover publications, 1945).

[9] E. T. Jaynes, Physical Review **106**, 620 (1957).

[10] A. Caticha and A. Giffin, AIP Conf. Proc. **872**, 31 (2006).

[11] E. T. Jaynes, Ann. Rev. Phys. Chem. **31**, 579 (1980).

[12] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Reviews of Modern Physics **85**, 1115 (2013).

[13] D. González, S. Davis, and G. Gutiérrez, Found. Phys. **44**, 923 (2014).

[14] S. Davis and D. González, J. Phys. A: Math. Theor. **48**, 425003 (2015).

[15] M. J. Hazoglou, V. Walther, P. D. Dixit, and K. A. Dill, J. Chem. Phys. **143**, 051104 (2015).

[16] H. Wan, G. Zhou, and V. A. Voeltz, J. Chem. Theory Comput. **12**, 5768 (2016).

[17] C. Cafaro and S. A. Ali, Phys. Rev. E **94**, 052145 (2016).

[18] P. Dixit, J. Wagoner, C. Weistuch, S. Pressé, K. Ghosh, and K. A. Dill, J. Chem. Phys. **148**, 010901 (2018).