Hamiltonian formalism and path entropy maximization

Sergio Davis and Diego González

Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile

E-mail: sdavis@gnm.cl and dgonzalez@gnm.cl

Received 15 May 2015, revised 10 August 2015
Accepted for publication 27 August 2015
Published 21 September 2015

Abstract
Maximization of the path information entropy is a clear prescription for constructing models in non-equilibrium statistical mechanics. Here it is shown that, following this prescription under the assumption of arbitrary instantaneous constraints on position and velocity, a Lagrangian emerges which determines the most probable trajectory. Deviations from the probability maximum can be consistently described as slices in time by a Hamiltonian, according to a nonlinear Langevin equation and its associated Fokker–Planck equation. The connections unveiled between the maximization of path entropy and the Langevin/Fokker–Planck equations imply that missing information about the phase space coordinate never decreases in time, a purely information-theoretical version of the second law of thermodynamics. All of these results are independent of any physical assumptions, and thus valid for any generalized coordinate as a function of time, or any other parameter. This reinforces the view that the second law is a fundamental property of plausible inference.

Keywords: path entropy, Langevin, Fokker–Planck, maximum caliber

1. Introduction

Jaynes’ principle of path entropy maximization (also known as the maximum caliber principle, or MaxCal for short) [1] is a clear prescription for the construction of dynamical models, both in non-equilibrium statistical mechanics [2, 3] as well as for any dynamical process [4, 5].

In this work, we derive consequences of the validity of the maximum caliber principle for the problem of estimating the trajectory of a coordinate, for instance, the position of a particle.
as a function of time. This framework will allow us to describe either classical mechanical
systems under uncertainty (e.g. under the influence of random forces) or stochastic signals
such as time series. Position and time are familiar concepts to us, but the results of this work
apply for inferences about any quantity \( q(s) \) parameterized by a continuous index \( s \), for
instance, in geometrical problems under uncertainty.

In the context of classical mechanics it is unavoidable to connect the maximum caliber
formalism, in which the probability of paths is proportional to the exponential of an action,
\[ P[x] \propto \exp(A[x]), \]
with Feynman’s path integral formalism in quantum mechanics \[6\].
A related attempt to bridge path integrals and classical mechanics is found in \[7\], although in
this case the probability distribution is explicitly constructed to suppress non-classical paths.
Connections between thermodynamics and dynamical systems (although without invoking
the idea of probabilities of paths) is found in \[8\] and \[9\].

The aim of this work is twofold: in the first place, it is interesting to explore to what
extent the structure of the dynamical framework of classical (in the sense of non-quantum)
physics is already contained in the simple idea of maximization of path entropy. However, the
results of this work may also result in powerful tools applicable to the continuous maximum
 caliber formalism.

2. The maximum caliber principle

Suppose for an unknown coordinate \( x \) described as a function of time \( t \) we only know the
expectation of a function \( f(x, \dot{x}; t) \) (over the distribution of possible trajectories \( x(t) \)) as a
function of time \( F(t) \), that is,
\[ \langle f(x, \dot{x}; t) \rangle_I = F(t), \] (1)
for every instant in the interval \([t_i, t_f]\). According to the maximum caliber principle, the
optimal assignment of probability for the trajectories maximizes
\[ S = -\int Dx(0)P[x(0)|I] \ln \frac{P[x(0)|I]}{P[x(0)|I_0]}, \] (2)
where \( I \) denotes all the given information about the problem, in particular the constraint in
equation (1), and \( I_0 \) is the complete ignorance state. The path probability \( P[x(0)|I_0] \) is the
\textit{a priori} measure needed for a consistent definition of entropy in continuous systems.
Imposing this constraint we can write the resulting probability as
\[ P[x(0)|I] = \frac{1}{Z[\lambda(0)]} P[x(0)|I_0] \exp \left( -\int_{t_i}^{t_f} dr \lambda(r) f(x(r), \dot{x}(r); t) \right). \] (3)
where \( \lambda \) is a Lagrange multiplier function. In the following we will assume, for simplicity of
the analysis, that the prior measure \( P[x(0)|I_0] \) is flat. If this is not the case, it can be absorbed
as an extra term in the action \( A[x(0)] \), to be defined below.

If we define a Lagrangian
\[ L(x, \dot{x}; t) = \alpha \lambda(t) f(x, \dot{x}; t), \] (4)
we can rewrite equation (3) as
\[ P[x(0)|I] = \frac{1}{Z[\lambda]} \exp \left( -\frac{1}{\alpha} A[x(0)] \right), \] (5)
where the functional $A$ is the action, defined as
\[
A[x(t)] = \int_{t_i}^{t_f} dt L(x(t), \dot{x}(t); t),
\]
and the constant $\alpha$, with dimensions of action, is extracted only to ensure the exponent in equation (5) is adimensional. Equation (5) is analogous to the probability amplitude assigned to a trajectory in the Feynman path integral formalism [6] (except the exponent here is real-valued), with $\alpha$ being a constant analogous to Planck’s constant. We see that the most probable path is automatically prescribed by the principle of minimum action, and thus is a solution of the Euler–Lagrange equation
\[
\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}.
\]

It is also immediately clear from this formalism that geometric constraints of the form $g_k(x(t); t) = 0$ will only add terms to the Lagrangian, as prescribed by classical mechanics [10],
\[
L(x, \dot{x}; t) = \alpha \lambda(t) f(x, \dot{x}; t) + \sum_k \mu_k(t) g_k(x(t); t)
\]
\[
= \alpha \lambda(t) \left[ f(x, \dot{x}; t) + \sum_k \tilde{\mu}_k(t) g_k(x(t); t) \right],
\]
where the $\tilde{\mu}_k = \mu_k/(\alpha \lambda)$ are (rescaled) Lagrange multipliers. Thus, from the point of view of maximum caliber, there is no conceptual distinction between the original Lagrangian and the terms from holonomic constraints: the entire Lagrangian always arises from constraints.

A striking example of this loss of distinction occurs when the quantity $F$ which is known and imposed as a constraint is the instantaneous joint probability $P(x(t), \dot{x}(t)|I)$. This constraint can be written as
\[
\left\{ \delta(x(t) - x)\delta(\dot{x}(t) - \dot{x}) \right\} I = P(x(t), \dot{x}(t)|I),
\]
i.e., the constraining function $f(x, \dot{x}; t)$ is the product of delta functions. If the probability in the right hand side is known for all $x$ and $\dot{x}$ in the state space $(x, \dot{x})$ where the probability is non-zero, the Lagrange multiplier function will now depend on $(x, \dot{x}, t)$ and the (normalized) action is
\[
\frac{A}{\alpha} = \int_{t_i}^{t_f} dt \int dx d\dot{x} \lambda(x, \dot{x}; t) \delta(x(t) - x) \delta(\dot{x}(t) - \dot{x})
\]
\[
= \int_{t_i}^{t_f} dt \lambda(x(t), \dot{x}(t); t),
\]
i.e. the Lagrange multiplier function itself becomes the Lagrangian. This means, although in principle the Lagrangian arising from equation (4) is arbitrary (determined by the choice of constraining function), a particular Lagrangian is singled out, the Lagrange multiplier of the probability.

3. Expectation of functionals

In order to obtain additional relations for expectation values of arbitrary functionals over the distribution given by equation (3), we apply the finite difference method [11] and map the continuous dynamical problem into a maximum entropy problem with finite number of
degrees of freedom by discretizing time, that is, replacing the continuous trajectory $x(t)$ by a vector of $N$ components, $\vec{x} = (x_1, \ldots, x_N)$, equation (3) then reduces to a maximum entropy solution

$$P(\vec{x} | I) = \frac{1}{Z} \exp\left(-\frac{1}{\alpha}A(\vec{x})\right)$$

with the action now replaced by a scalar field $A(\vec{x})$, given by

$$A(\vec{x}) = \Delta t \sum_{j=1}^{N} L_j (x_j, (x_{j+1} - x_j)/\Delta t).$$

There is a connection between expectation values provided by the conjugate variables theorem \cite{12},

$$\left\langle \nabla \cdot \psi(\vec{x}) \right\rangle_I = \frac{1}{\alpha} \left\langle \psi \cdot \nabla A(\vec{x}) \right\rangle_I.$$  

Let us choose $\vec{v} = \hat{e}_k W(\vec{x})$ where $\hat{e}_k$ the unit vector in along the $k$th coordinate and the trial field $W$ is of the form

$$W(\vec{x}) = \Delta t \sum_{j=1}^{N} \omega_j (x_j, (x_{j+1} - x_j)/\Delta t).$$

Then,

$$\left\langle \frac{\partial}{\partial x_k} W(\vec{x}) \right\rangle_I = \frac{1}{\alpha} \left\langle W \frac{\partial}{\partial x_k} A(\vec{x}) \right\rangle_I.$$  

Now, for both functionals $W(\vec{x})$ and $A(\vec{x})$ we have \cite{11},

$$\frac{\partial}{\partial \delta t \partial x_k} \rightarrow \frac{\delta}{\delta x(t)}.$$  

and the functional version of the conjugate variables theorem is given by

$$\left\langle \frac{\delta W}{\delta x(t)} \right\rangle_I = \frac{1}{\alpha} \left\langle W[x(t)] \frac{\delta A}{\delta x(t)} \right\rangle_I,$$

with $W[x(t)]$ a trial functional with the form

$$W[x()] = \int_{t_i}^{t_f} dt \omega(x(t), \dot{x}(t); t).$$

This is the equivalent of the identity (7.30) in \cite{6} for the quantum mechanical path integrals.

By explicitly replacing the Lagrangian $L$ and the trial function $\omega$, we can compactly write equation (17) as

$$\langle \hat{E}_i \omega \rangle_I = \frac{1}{\alpha} \langle W[x()] \hat{E}_i L \rangle_I,$$

where we have introduced the operator $\hat{E}_i$ as

$$\hat{E}_i G(x, \dot{x}; t) = \left( \frac{\partial}{\partial x} - \frac{d}{dt} \frac{\partial}{\partial \dot{x}} \right) G(x, \dot{x}; t).$$
Choosing $W[x(t)] = 1$ tells us that
\[ \langle \dot{E}_t, L \rangle_t = 0 \] (21)
for all instants $t$, i.e., the Euler–Lagrange equation is valid in expectation over the ensemble of trajectories, not only for the most probable one. This suggests treating $\dot{E}_t$ itself as a random variable with zero expectation for the purpose of constructing a Langevin equation. However, we need to establish the predicted correlation $\langle \dot{E}_t, \dot{E}_t \rangle_t$ of the quantity $\dot{E}_t$. Now we use $\dot{E}_t$ in equation (19) to show that
\[ \langle \dot{E}_t, \dot{E}_t \rangle_t = \left( \frac{\delta^2 A}{\delta x(t) \delta x(t')} \right)_t = \frac{1}{\alpha} \left( \frac{\delta^2 A}{\delta x(t) \delta x(t')} \right)_t. \] (22)

As the action has the form in equation (6), its second functional derivative is
\[ \frac{\delta^2 A}{\delta x(t) \delta x(t')} = \frac{\delta}{\delta x(t')} \left( \frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right) \left| t \right. \]
\[ = \frac{\partial}{\partial x'} \left( \frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right) - \frac{d}{dt'} \frac{\partial}{\partial x'} \left( \frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right) \] (23)
and this is only non-zero for $t = t'$, thus in all generality we can write it as
\[ \langle \dot{E}_t, \dot{E}_t \rangle_t = 2R(t) \delta(t - t') \] (24)
with $R(t)$ a function of time to be determined. Here we note that, by taking the limit $t \to t'$ in equation (24), it follows that $R(t)$ cannot be negative. From equations (21) and (24), we now formally write the stochastic differential equation (SDE),
\[ \dot{E}_t = \frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} = \sqrt{2R(t)} \xi(t) \] (25)
with $\xi(t)$ an unbiased, uncorrelated ‘noise’,
\[ \langle \xi(t) \rangle_t = 0 \] (26)
\[ \langle \xi(t) \xi(t') \rangle_t = \delta(t - t') \] (27).

It is crucial to note at this point that $R(t)$ does not necessarily represent a ‘random’ physical influence: as the probability interpretation when using the maximum entropy and maximum caliber principles is unavoidably Bayesian, $R(t)$ represents inaccessible or hidden information. Without this uncertainty, the system follows the trajectory which minimizes the action (equation (25) reduces to the Euler–Lagrange equation, equation (7)).

Equation (25) is, in general, a second-order SDE, but we can obtain a system of two first-order SDEs by expressing it in canonical form [11]. For this we define the Hamiltonian $\mathcal{H}$ corresponding to the Lagrangian $L$ as the Legendre transformation
\[ \mathcal{H}(x, p) = p\dot{x} - L(x, \dot{x}; t) = \alpha\lambda(t) \left( \frac{\partial f}{\partial \dot{x}} - f \right) \bigg|_{x, p}, \] (28)
together with the momentum $p = \partial L/\partial \dot{x}$. Now we have defined a phase space $(x, p)$ and by virtue of the Legendre transformation the following relations hold, independently of the form of the Lagrangian,
\[ \frac{\partial \mathcal{H}}{\partial x} = - \frac{\partial L}{\partial x}, \] (29)
By introducing the Poisson brackets
\[ \{ F, G \} = \frac{\partial F}{\partial x} \frac{\partial G}{\partial p} - \frac{\partial F}{\partial p} \frac{\partial G}{\partial x} \]
we see that the required system of first-order SDEs is a nonlinear Langevin equation in phase space,
\[ \dot{\vec{r}} = \left\{ \vec{r}, \mathcal{H} \right\} + R(t) \xi(t), \]
where we have used the Einstein summation convention, \( \vec{r} = (x, p) \) and the only non-zero component of the uncertainty matrix \( R_{ij} \) is
\[ R_{\alpha\beta} = \lim_{t \to 0} \frac{1}{2e} \mathbb{E}\left[ \Delta \vec{r}_{\alpha}(t + \epsilon) \Delta \vec{r}_{\beta}(t) \right] = R_{\alpha\beta}(t), \]
where \( \Delta \vec{r}_{\alpha} = \vec{r}_{\alpha}(t + \epsilon) - \vec{r}_{\alpha}(t) \). Therefore we can analyze the time evolution of the phase space probability \( P \) by solving
\[ \frac{\partial P}{\partial t} + \{ P, \mathcal{H} \} = R(t) \frac{\partial^2 P}{\partial p^2}. \]

This is a Boltzmann transport equation with collision term proportional to the uncertainty \( R(t) \). Note that, for an arbitrary Hamiltonian, in the case of complete certainty \( R(T) \to 0 \) we recover the Liouville equation. It is useful to write equation (37) in logarithmic form, namely,
\[ \frac{\partial}{\partial t} \ln P + \langle \ln P, \mathcal{H} \rangle_t = R(t) \left[ \frac{\partial^2}{\partial p^2} \ln P + \left( \frac{\partial}{\partial p} \ln P \right)^2 \right]. \quad (38) \]

Multiplying by an arbitrary function \( G(x, p; t) \) and taking expectation over \( P \) we have

\[ \langle G \frac{\partial}{\partial t} \ln P \rangle_t + \langle G \{ \ln P, \mathcal{H} \} \rangle_t = R(t) \langle G \left[ \frac{\partial^2}{\partial p^2} \ln P + \left( \frac{\partial}{\partial p} \ln P \right)^2 \right] \rangle_t, \quad (39) \]

**FIXME**

Here we apply the general ‘fluctuation-dissipation’ relation for an arbitrary parameter \( \alpha \) of a distribution \( P \),

\[ \frac{\partial}{\partial \alpha} \langle G \rangle_{t, \alpha} = \left\langle \frac{\partial G}{\partial \alpha} \right\rangle_{t, \alpha} + \left\langle G \frac{\partial}{\partial \alpha} \ln P \right\rangle_{t, \alpha}, \quad (40) \]

with \( \alpha = t \), and the conjugate variables theorem [12], as

\[ \left\langle \frac{\partial G}{\partial \alpha} \right\rangle_t = - \left\langle G \frac{\partial}{\partial \alpha} \ln P \right\rangle_t. \quad (41) \]

After some algebraic manipulations and repeated use of equation (40) and (41) to eliminate \( \ln P \), we obtain a classical (non-quantum) analog of the Ehrenfest theorem [15],

\[ \frac{d}{dt} \langle G \rangle_{t, t} = \left\langle \frac{\partial G}{\partial t} \right\rangle_{t, t} + \langle \{ G, \mathcal{H} \} \rangle_{t, t} + R(t) \langle \frac{\partial^2 G}{\partial p^2} \rangle_{t, t}, \quad (42) \]

which gives us the time evolution of an arbitrary phase space function \( G \), and also provides us with a direct way to compute \( R(t) \) from time-dependent averages.

An important remark we address here is the fact that the expectation values over trajectories are completely described by the Lagrangian (equation (19)), while the instantaneous properties, representing the system as successive ‘slices’ in time, are in turn described by the Hamiltonian (equation (42)).

### 5. Second law of thermodynamics

Replacing \( G = \ln P \) in equation (42) and taking into account that, because of equation (40) and equation (41), \( \langle \partial_t \ln P \rangle_{t, t} = 0 \) and \( \langle \{ \ln P, \mathcal{H} \} \rangle_{t, t} = 0 \), we see that

\[ \frac{dS}{dt} = R(t) \left\langle \left( \frac{\partial}{\partial p} \ln P \right)^2 \right\rangle_{t, t} \geq 0, \quad (43) \]

with \( S(t) = -\langle \ln P \rangle_{t,t} \) the information entropy of the distribution \( P \) at time \( t \). Therefore, missing information about the location in phase space never decreases with time, and its increase is solely due to the uncertainty. In other words, irreversibility is equivalent to uncertainty over the trajectory followed by the system, whatever its origin (e.g. sensibility to initial conditions, external driving).
6. Concluding remarks

In summary, we have explored the consequences of constructing a maximum caliber dynamical model using only instantaneous information about position and velocity. In the first place, the problem of prediction of the most probable trajectory reduces to standard classical mechanics, complete with a Lagrangian, a phase space and a Hamiltonian. For the description of deviations around the most probable trajectory, the formalism naturally leads to a nonlinear Langevin equation with uncorrelated noise in phase space, and its corresponding Fokker–Planck equation. This Fokker–Planck equation is completely determined by the Hamiltonian of the problem.

Given the striking resemblance of equations (5), (17) and (42) to real-valued versions of the Feynman path integral formulation of quantum mechanics and the Ehrenfest theorem, respectively, it remains to be seen if the idea of path entropy maximization could lead to their quantum mechanical (complex-valued) versions under suitable constraints, as well as its connection with Nelson’s stochastic mechanics [16, 17].

Acknowledgments

The authors gratefully acknowledge funding from FONDECYT grant number 1140514. DG also acknowledges funding from CONICYT PhD fellowship 21140914.

References

[1] Jaynes E T 1980 Ann. Rev. Phys. Chem. 31 579
[2] Dewar R 2003 J. Phys.: Math. Gen. 36 631
[3] Grandy W T 2008 Entropy and the Time Evolution of Macroscopic Systems (Oxford: Oxford University Press)
[4] Haken H 1986 Z. Phys. B 63 505
[5] Pressé S, Ghosh K, Lee J and Dill K A 2013 Rev. Mod. Phys. 85 1115
[6] Feynman R P and Hibbs A R 2005 Quantum Mechanics and Path Integrals (New York: McGraw-Hill)
[7] Gozzi E, Reuter M and Thacker W D 1989 Phys. Rev. D 40 3363
[8] Grmela M 2013 Comput. Math. Appl. 65 1457
[9] Grmela M 2014 Entropy 16 1652
[10] Lanzcos C 1970 The Variational Principles of Mechanics (New York: Dover)
[11] Gelfand I M and Fomin S V 2000 Calculus of Variations (New York: Dover)
[12] Davis S and Gutierrez G 2012 Phys. Rev. E 86 051136
[13] Yura Y, Takayasu H, Sornette D and Takayasu M 2014 Phys. Rev. Lett. 112 098703
[14] Risken H 1996 The Fokker–Planck Equation: Methods of Solution and Applications (Berlin: Springer)
[15] Shankar R 1994 Principles of Quantum Mechanics (New York: Plenum)
[16] Nelson E 1966 Phys. Rev. 150 1079
[17] González D, Davis S and Gutierrez G 2014 Found. Phys. 44 923