# Newtonian Dynamics from the Principle of Maximum Caliber

**Diego González · Sergio Davis ·
Gonzalo Gutiérrez**

**Abstract** The foundations of Statistical Mechanics can be recovered almost in their entirety from the principle of maximum entropy. In this work we show that its non-equilibrium generalization, the principle of maximum caliber (Jaynes, Phys Rev 106:620–630, 1957), when applied to the unknown trajectory followed by a particle, leads to Newton's second law under two quite intuitive assumptions (both the expected square displacement in one step and the spatial probability distribution of the particle are known at all times). Our derivation explicitly highlights the role of mass as an emergent measure of the fluctuations in velocity (inertia) and the origin of potential energy as a manifestation of spatial correlations. According to our findings, the application of Newton's equations is not limited to mechanical systems, and therefore could be used in modelling ecological, financial and biological systems, among others.

## 1 Introduction

Newtonian mechanics is a cornerstone of our civilization. From its simpler formulation, $F = ma$, to Lagrangian and Hamiltonian mechanics it has evolved an elegant mathematical apparatus around it. However, we still have to assume the principle of

D. González (✉) · S. Davis · G. Gutiérrez
Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias,
Universidad de Chile, Casilla 653, Santiago, Chile
e-mail: dgonzalez@gnm.cl

S. Davis
e-mail: sdavis@gnm.cl

G. Gutiérrez
e-mail: gonzalo@macul.ciencias.uchile.cl

minimum action as a postulate (that is, if we do not consider Newtonian mechanics as a limiting consequence of quantum mechanics).

There have been attempts to derive Newtonian dynamics from different principles. Particularly related to this work, it has been previously derived from information-geometric arguments leading to the idea of entropic dynamics [1]. This idea is based on the assumption of an irreducible uncertainty in the position of a particle, implying an information metric for space from which Newton's second law naturally emerges. Caticha and Cafaro's derivation is founded on Jaynes' maximum entropy principle [2] (MaxEnt for short), a generic rule for the construction of probabilistic models.

In 1957, Edwin T. Jaynes [2] postulated that Statistical Mechanics has to be understood, not as a physical theory in the same footing as, say, classical mechanics or electromagnetism, but as an application of statistical inference to a system with macroscopically large numbers of degrees of freedom. The question was reversed from "given the microscopic evolution of the system, what is the probability distribution for the macroscopic quantities?" to "given a few known macroscopic properties, what are the possible microstates compatible with said knowledge?". The answer, as initially proposed by Gibbs, was the probability distribution with maximum entropy $S = -\sum_i P_i \ln P_i$ subjected to constraints reflecting the known macroscopic properties. Jaynes, after the work of Shannon in information theory, realized that this procedure (maximization of $S$ constrained only by the known information) is not limited to Statistical Mechanics but a valid principle in any problem of statistical inference. This is the principle of Maximum Entropy. Due to the uniqueness of Shannon's entropy in characterizing uncertainty, it is the most unbiased procedure for the construction of statistical models. Later, it has been axiomatically derived [3,4] from requirements of internal consistency, so its validity is independent of any meaning assigned to the quantity $S$.

A full understanding of the maximum entropy principle requires the adoption of the Bayesian viewpoint on probability: An "absolute" probability distribution $P(x)$ does not exist, only conditional probability distributions $P(x|H)$ where $H$ represents a given state of knowledge. For a given $H$, MaxEnt gives us the most unbiased model we can propose with our (possibly limited) information, therefore we cannot guarantee beforehand that it is the correct model. As we acquire more and more information, $P(x|H)$ becomes increasingly more independent of $H$, and for all practical purposes the distribution becomes objective (independent of the observer).

The principle of Maximum Caliber [5,6] generalizes this idea to dynamical systems, including time explicitly. It has been applied recently to discrete dynamics [7] and earlier to derive the Fokker–Planck equations [8] and the Markov process formalism [9].

In this work we show that if we use the maximum caliber principle to find the unknown trajectory of a particle, there are two general conditions that automatically lead to Newton's second law, namely that (a) the expected square displacement per step is known at all times, and (b) that the time-independent probability of finding the particle at any coordinate is also known. Knowledge of both (a) and (b) leads to Newton's second law in expectation over trajectories, and what is perhaps more interesting, any dynamical system not following Newton's second law has to violate at least one of these assumptions.

## 2 The Maximum Entropy and Maximum Caliber Formalism

Consider a system with $N$ degrees of freedom, whose states are denoted by vectors $\vec{x} = (x_1, \ldots, x_N)$. Suppose the expectation values of $M$ functions $f_i(\vec{x})$ are known. This represents a *Bayesian state of knowledge*, that we will denote by $H$. Maximization of the Shannon entropy

$$S = -\int d\vec{x}\, P(\vec{x}|H) \ln P(\vec{x}|H) \tag{1}$$

under the constraints $\langle f_i(\vec{x}) \rangle_H = F_i$ leads to the MaxEnt model

$$P(\vec{x}|H) = \frac{1}{Z(\vec{\lambda})} \exp\left(-\sum_{k=1}^{M} \lambda_k f_k(\vec{x})\right), \tag{2}$$

where the value of the Lagrange multipliers $\lambda_k$ needed to impose the $M$ constraints can be determined from

$$-\frac{\partial}{\partial \lambda_k} \ln Z(\vec{\lambda}) = F_k. \tag{3}$$

Equation 3, being a nonlinear equation is usually impractical to solve. Moreover, it needs the partition function explicitly. However, it has been recently shown [10] that for the Lagrange multipliers the equality

$$\left\langle \nabla \cdot \vec{v} \right\rangle = \sum_{k=1}^{M} \lambda_k \left\langle \vec{v} \cdot \nabla f_k \right\rangle \tag{4}$$

holds, with $\vec{v}$ an arbitrary differentiable vector field, and this provides a linear system of equations for $\vec{\lambda}$. The most probable microstate $\vec{x}_0$ is such that

$$\nabla P(\vec{x}|H)|_{\vec{x}=\vec{x}_0} = 0,$$

therefore the function

$$G(\vec{x}) = \sum_{k=1}^{M} \lambda_k f_k(\vec{x})$$

which is a combination of all the constraints imposed, is also an extremum at $\vec{x}_0$.

In the case of dynamical systems, we now ask for the possible microscopical trajectories compatible with some known information $H$ (expressed as expectation values). The result is the probability distribution of trajectories $P[x(t)|H]$ which maximizes the Shannon entropy

$$\mathcal{S} = -\int \mathcal{D}x(t)\, P[x(t)|H] \ln P[x(t)|H] \tag{5}$$

given $H$. Note that $S$ is now defined as a functional integral, as indicated by the notation $\mathcal{D}x(t)$. This conveys the idea of integration over a space of trajectories $x(t)$, as is commonly denoted in Feynman's path integral formulation of quantum mechanics [11].

The probability distribution functional for the different possible trajectories is

$$P[x(t)|H] = \frac{1}{Z[\lambda(t)]} e^{-\int dt \lambda(t) f[x(t);t]}, \tag{6}$$

where, similarly to Eq. 3, the Lagrange multiplier function can be obtained from

$$-\frac{\delta}{\delta\lambda(t)} \ln Z[\lambda(t)] = \langle f[x(t);t] \rangle. \tag{7}$$

Here the most probable trajectory extremizes a particular functional (analogous to an "action" in classical mechanics)

$$A[x(t)] = \int dt \lambda(t) f[x(t);t] \tag{8}$$

depending on the constraining function. This leads to the question: without introducing the Lagrangian of classical mechanics explicitly, could it *emerge* naturally from simpler constraints in a Maximum Caliber problem?

## 3 Derivation of Newton's Second Law

Consider a single particle following an unknown trajectory $x(t)$ in one spatial dimension. This can be easily generalized to many particles in arbitrary dimensions, at the cost of overcomplicated notation.

If we discretize time, such that the continuous variable $t$ is sampled at intervals $\Delta t$ and becomes $t_k = k\Delta t$ (with $k = 0, \ldots, n-1$), the positions $x(t)$ have to be replaced with $x_k = x(t_k)$. Now the trajectory $x(t)$ itself becomes a vector $\vec{x} = (x_0, \ldots, x_{n-1})$, and the Lagrange function $\lambda(t)$ also becomes a vector $\vec{\lambda} = (\lambda_0, \ldots, \lambda_{n-1})$. This is the method of finite differences [12]. We can see that, using this discretized version, we recover Eqs. 2 and 3, therefore we can also employ the identity given by Eq. 4 for this kind of discretized Maximum Caliber problem.

We now impose the following constraints (expectations are to be interpreted over all possible trajectories)

$$\left\langle (x_i - x_{i-1})^2 \right\rangle = (\Delta t)^2 d_i{}^2 \tag{9}$$

$$\left\langle \delta(x_i - X) \right\rangle = P(x_i = X|H), \tag{10}$$

for all values of $i$ and $X$. The first constraint recognizes the fact that the expected square displacement in one (possible infinitesimal) step is known for all times, and

is equal to an arbitrary function $d_i{}^2$ times the time step. We expressed it in this form so that $d_i$ can remain finite when taking the limit $\Delta t \to 0$. The second constraint imposes that the static, time-independent probability distribution for the coordinate $x$ is also known.

Here we must remark that, as we are considering the Bayesian view of probability, the last constraint can always be fulfilled: $P(x_i = X|H)$ is a model comprising our knowledge of $x_i$ *under the state H*, not a physical property of the phenomenon. Different states of information $H$ may assign different probability distributions $P(x_i = X|H)$, and the "quality" of this information will be reflected in the resulting potential energy function emerging from this formalism (as will be seen later).

The probability distribution function for $\vec{x}$ is

$$P(\vec{x}|H) = \frac{1}{Z(\vec{\lambda})} \exp\left(-\sum_{i=0}^{n-1} \frac{\lambda_i}{(\Delta t)^2}(x_i - x_{i-1})^2 - \sum_{i=0}^{n-1} \int dX \mu(X)\delta(x_i - X)\right), \quad (11)$$

where $\lambda_i$ and $\mu(X)$ are Lagrange multipliers associated to the constraints in Eqs. 9 and 10, respectively.

After integrating the Dirac delta function, Eq. 11 becomes

$$P(\vec{x}|H) = \frac{1}{Z(\vec{\lambda})} \exp\left(-\sum_{i=0}^{n-1} \frac{\lambda_i}{(\Delta t)^2}(x_i - x_{i-1})^2 - \sum_{i=0}^{n-1} \mu(x_i)\right). \quad (12)$$

This is the probability of the particle taking a well-defined discretized trajectory $\vec{x}$, and is precisely the solution of a Maximum Entropy problem with $n$ degrees of freedom and $n$ Lagrange multipliers $\lambda_i$ (plus the function $\mu$).

The most probable trajectory for the particle follows a minimum action principle. Indeed, if we define $m_k = 2\lambda_k$ and $\Phi(x) = -\mu(x)$, we recover, from Eq. 12,

$$P(\vec{x}|H) = \frac{1}{Z} \exp\left(-A(\vec{x})\right) \quad (13)$$

where a discretized version of the classical action

$$A(\vec{x}) = \sum_{i=0}^{n-1} \left[\frac{1}{2}m_i \dot{x}_i^2 - \Phi(x_i)\right] \quad (14)$$

appears in the exponential. Clearly, Eqs. 13 and 14 in the continuum limit become

$$P[x(t)|H] = \frac{1}{Z} \exp\left(-A[x(t)]\right) \quad (15)$$

$$A[x(t)] = \int dt \mathcal{L}(x(t), \dot{x}(t); t),$$

and it follows that the most probable continuous trajectory $x(t)$ is the one that extremizes the classical action $A[x(t)]$ with Lagrangian

$$\mathcal{L}(x, \dot{x}; t) = \frac{1}{2}m(t)\dot{x}^2 - \Phi(x) \tag{16}$$

and associated Hamiltonian

$$\mathcal{H}(x, p; t) = \frac{p^2}{2m(t)} + \Phi(x). \tag{17}$$

This is the formalism of Classical Mechanics for a particle with time-dependent mass $m(t)$ subjected to a potential energy $\Phi(x)$. Both $m$ and $\Phi$ have emerged from the Lagrange multipliers associated with the constraints on the expected square of the step and the probability distribution of the coordinate, respectively. Thus we can say the following: whenever the information about the expected square of the step is important, the particle acquires mass, and whenever the information about which regions are more probable in space becomes important, the particle is subjected to a potential energy.

As Eq. 12 is a MaxEnt solution, Eq. 4 holds as

$$\left\langle \nabla \cdot \vec{v}(\vec{x}) \right\rangle = \sum_{i=0}^{n-1} \frac{\lambda_i}{(\Delta t)^2}\left\langle \vec{v}(\vec{x}) \cdot \nabla(x_i - x_{i-1})^2 \right\rangle + \sum_{i=0}^{n-1}\left\langle \vec{v}(\vec{x}) \cdot \nabla\mu(x_i) \right\rangle, \tag{18}$$

with $\vec{v}$ an arbitrary differentiable vector field, of our choosing. If we choose $\vec{v}$ such that it has a single component $k$, i.e. $v_i = \delta_{i,k}\omega(\vec{x})$ with $\omega$ an arbitrary scalar field, we obtain

$$\left\langle \frac{\partial \omega}{\partial x_k} \right\rangle = \sum_{i=0}^{n-1} \frac{\lambda_i}{(\Delta t)^2}\left\langle \omega(\vec{x}) \cdot 2(x_i - x_{i-1})(\delta_{i,k} - \delta_{i-1,k}) \right\rangle + \sum_{i=0}^{n-1}\left\langle \omega(\vec{x}) \cdot \mu'(x_i)\delta_{i,k} \right\rangle$$
$$= \frac{1}{(\Delta t)^2}\left\langle 2\omega(\vec{x})\Big[\lambda_k(x_k - x_{k-1}) - \lambda_{k+1}(x_{k+1} - x_k)\Big] \right\rangle + \left\langle \omega(\vec{x})\mu'(x_k) \right\rangle. \tag{19}$$

But recalling that the discrete forward derivative is

$$\dot{a}_i \approx \frac{a_{i+1} - a_i}{\Delta t}, \tag{20}$$

we can re-write Eq. 19 as

$$\left\langle \frac{\partial \omega}{\partial x_k} \right\rangle = -\left\langle \omega\Big(\dot{p}_k + \mu'(x_k)\Big) \right\rangle, \tag{21}$$

where $p_k = m_k\dot{x}_k$.

Considering $\omega = 1$ we finally obtain

$$\left\langle \dot{p}_k \right\rangle = -\left\langle \Phi'(x_k) \right\rangle. \tag{22}$$

which, in the continuous limit ($\Delta t \to 0$), becomes

$$\left\langle \dot{p}(t) \right\rangle = -\left\langle \Phi'(x(t)) \right\rangle, \tag{23}$$

the expectation value of Newton's second law with momentum $p = m\dot{x}$ and potential energy $\Phi(x)$.

In appendix A we explore the validity of some aspects of the canonical formalism, namely the Poisson bracket, for the expectation over trajectories.

## 4 Concluding Remarks

We have found that two simple constraints are sufficient to recover Newton's second law in expectation for the probable trajectories of a particle. The first constraint, on the step size as a function of time, leads to the existence of an inertial mass $m(t)$ proportional to the Lagrange multiplier $\lambda(t)$. To understand the meaning of this, remember that for any variational problem solved using Lagrange multipliers, the larger the value of the multiplier, the more restrictive (and therefore more relevant) the constraint. An irrelevant constraint has always a vanishing multiplier. As Jaynes [13] (p. 945) clearly states, "The Lagrange multipliers $\lambda_k$ in the MAXENT formalism have therefore a deep meaning: $\lambda_k$ is the 'potential' of the datum $R'_k$, that measures how important a constraint it represents."

Now we motivate the following principle: constraints related to conserved quantities are always more relevant. For instance, this explains the fact that the canonical ensemble in equilibrium statistical mechanics is correctly derived just from a single constraint, the energy or expectation of the Hamiltonian, which is an integral of motion. Another illustration is the following: suppose we are trying to recover the trajectory of a particle from information about the distance to a particular point. If this distance is a constant, this is enough to isolate a unique trajectory, the circle. If we only know that the distance varies between $r_1$ and $r_2$, the number of compatible trajectories will increase with $\Delta r = r_2 - r_1$, thus the strength of the constraint will correspondingly decrease with increasing $\Delta r$.

Given the earlier discussion, the closer $d_i^2$ is to be a conserved quantity, the more relevant the first constraint is. In this case, $\lambda(t)$ is large and therefore, $m(t)$ is also large. Conversely, if the value of $m$ is small, this means $\lambda(t)$ is small and therefore $d_i^2$ has larger fluctuations. In the continuous limit it is the instantaneous speed that fluctuates (there is a non-zero acceleration). This embodies the idea of inertia, and is reminiscent of the ideas of Smolin [14] and of Nelson [15] about inertia being inversely proportional to the size of quantum fluctuations.

## Appendix 1: Canonical Coordinates and Poisson Brackets

An interesting question is how much of the formalism of classical mechanics we can recover from Eq. 22. The fact that most of the structure of classical mechanics is contained in the definition and properties of the Poisson bracket, motivates us to search for an operation analogous to this bracket under the Maximum Caliber formalism.

For arbitrary differentiable functions $f(x, p)$ and $g(x, p)$ the Poisson bracket is defined as

$$\{f, g\} = \frac{\partial f}{\partial x}\frac{\partial g}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial g}{\partial x}, \tag{24}$$

and it is such that

$$\frac{df}{dt} - \frac{\partial f}{\partial t} = \{f, \mathcal{H}\} \tag{25}$$

holds. Let us compute the expectation of the left hand side,

$$\left\langle\frac{df}{dt}\right\rangle - \left\langle\frac{\partial f}{\partial t}\right\rangle = \left\langle\frac{\partial f}{\partial x_k}\dot{x}_k + \frac{\partial f}{\partial p_k}\dot{p}_k\right\rangle, \tag{26}$$

which using Eq. 21 with $\omega = \partial f/\partial p_k$ can be written as

$$\left\langle\frac{df}{dt}\right\rangle - \left\langle\frac{\partial f}{\partial t}\right\rangle = \left\langle\frac{\partial f}{\partial x_k}\dot{x}_k - \frac{\partial}{\partial x_k}\left(\frac{\partial f}{\partial p_k}\right) - \frac{\partial f}{\partial p_k}\Phi'(x_k)\right\rangle \tag{27}$$

Now using our classical Hamiltonian (Eq. 17) we recognize its derivatives

$$\dot{x}_k = \frac{\partial\mathcal{H}}{\partial p_k} \tag{28}$$

$$\Phi'(x_k) = \frac{\partial\mathcal{H}}{\partial x_k} \tag{29}$$

and, upon replacing, we have

$$\left\langle\frac{df}{dt}\right\rangle - \left\langle\frac{\partial f}{\partial t}\right\rangle = \left\langle\frac{\partial f}{\partial x_k}\frac{\partial\mathcal{H}}{\partial p_k} - \frac{\partial}{\partial x_k}\left(\frac{\partial f}{\partial p_k}\right) - \frac{\partial f}{\partial p_k}\frac{\partial\mathcal{H}}{\partial x_k}\right\rangle \tag{30}$$

leading finally to

$$\left\langle\frac{df}{dt}\right\rangle - \left\langle\frac{\partial f}{\partial t}\right\rangle = \left\langle\{f, \mathcal{H}\}\right\rangle - \left\langle\frac{\partial}{\partial x_k}\left(\frac{\partial f}{\partial p_k}\right)\right\rangle. \tag{31}$$

So, in expectation we find a Poisson bracket analog with an additional term. For the particular case $f = \mathcal{H}$, we obtain

$$\left\langle\frac{d\mathcal{H}}{dt}\right\rangle = -\left\langle\frac{\partial}{\partial x_k}\left(\frac{\partial\mathcal{H}}{\partial p_k}\right)\right\rangle, \tag{32}$$

which reduces to

$$s\left\langle \frac{d\mathcal{H}}{dt} \right\rangle = -\left\langle \frac{\partial \dot{x}_k}{\partial x_k} \right\rangle = 0, \tag{33}$$

using the centered difference,

$$\dot{a}_i \approx \frac{a_{i+1} - a_{i-1}}{2\Delta t}. \tag{34}$$

Therefore we have shown that, for a Hamiltonian with the form given in Eq. 17, the energy is conserved in expectation. Note that using one-sided differences is not appropriate here, as one obtains $\pm 1/\Delta t$ depending on forward or backward.

## References

1. Caticha, A., Cafaro, C.: From information geometry to newtonian dynamics. AIP Conf. Proc. **954**, 165 (2007)
2. Jaynes, E.T.: Information theory and statistical mechanics. Phys. Rev. **106**, 620–630 (1957)
3. J. E. Shore and R. W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. IEEE Trans. Info. Theory, IT. **26**, 26–37 (1980)
4. Skilling, J.: The axioms of maximum entropy. In: Erickson, G.J., Smith, C.R., (eds.) Maximum Entropy and Bayesian Methods in Science and Engineering, pp. 173–187. Kluwer Academic Publishers, New York (1988)
5. Jaynes, E.T.: The minimum entropy production principle. Ann. Rev. Phys. Chem. **31**, 579–601 (1980)
6. Press, S., Ghosh, K., Lee, J., Dill, K.A.: Principles of maximum entropy and maximum caliber in statistical physics. Rev. Mod. Phys. **85**, 1115 (2013)
7. Stock, G., Ghosh, K., Dill, K.A.: Maximum caliber: a variational approach applied to two-state dynamics. J. Chem. Phys. **128**, 194102 (2008)
8. Haken, H.: A new access to path integrals and fokker planck equations via the maximum calibre principle. Z. Phys. B. Cond. Matt. **63**, 505–510 (1986)
9. Ge, H., Press, S., Ghosh, K., Dill, K.: Markov processes follow from the principle of maximum caliber. J. Chem. Phys. **136**, 064108 (2012)
10. Davis, S., Gutiérrez, G.: Conjugate variables in continuous maximum-entropy inference. Phys. Rev. E **86**, 051136 (2012)
11. Feynman, R.P., Hibbs, A.R.: Quantum Mechanics and Path Integrals, Dover Books, (2005)
12. Gelfand, I.M., Fomin, S.V.: Calculus of Variations. Dover Publications, (2000)
13. Jaynes, E.T.: On the rationale of maximum-entropy methods. Proc. IEEE **10**, 939–952 (1982)
14. Smolin, Lee: Quantum fluctuations and inertia. Phys. Lett. **113A**, 408 (1986)
15. Nelson, E.: Derivation of the schrdinger equation from newtonian mechanics. Phys. Rev. **150**, 1079–1085 (1966)