Macroscopic time evolution and MaxEnt inference for closed systems with Hamiltonian dynamics

Domagoj Kuić,* Paško Županović,† and Davor Juretić‡
University of Split, Faculty of Science, N. Tesle 12, 21000 Split, Croatia
(Dated: November 18, 2010)

Abstract

MaxEnt inference algorithm from the information theory is relevant for the time evolution of macroscopic systems which represents a problem of incomplete information. Two different approaches using MaxEnt algorithm are considered in this work, both applied to the problem of predicting macroscopic time evolution of closed systems with Hamiltonian dynamics. The first approach is based on Liouville equation for the conditional probability distribution introduced as a strict microscopic constraint on time evolution in phase space. Phase space paths determined by solutions of Hamilton's equations are introduced as a representation of information about microscopic dynamics. Path probability distribution is defined for the set of phase space paths. The conditional probability distribution is defined for the set of microstates associated with the set of phase space paths. The MaxEnt inference algorithm with Shannon's concept of conditional information entropy is then used for prediction of time evolution consistently with this strict microscopic constraint on time evolution in phase space. It is concluded consequently that this approach is equivalent to complete knowledge of microscopic dynamics. The second approach is based on the same concepts with a difference that Liouville equation for the conditional probability distribution is introduced as a macroscopic constraint given by phase space average. In this approach we consider incomplete nature of our information about microscopic dynamics in a rational way that is consistent with Jaynes' formulation of predictive statistical mechanics and the concept of macroscopic reproducibility for time dependent processes. Maximization of the conditional information entropy subject to this macroscopic constraint leads to a loss of correlation between the initial phase space paths and final microstates. It is known from information theory that the information entropy is the upper bound on the conditional information entropy, with the upper bound attained only in case of complete loss of correlation. In reproducible time evolution of a closed system, information entropy of the microstate probability distribution has its maximum value consistent with constraints. This is a consequence of macroscopic reproducibility and the information loss about possible but otherwise unknown microstates into which the system finds itself during the reproducible process. In this alternative approach to prediction of macroscopic time evolution, maximization of the conditional information entropy includes all described properties and is equivalent to the loss of statistical correlation. Irreversibility appears as a consequence of gradual loss of information about possible microstates of the system, as was suggested by Jaynes.

PACS numbers: 02.50.Tt, 05.20.Gg, 05.70.Ln, 45.20.Jj
Keywords: MaxEnt algorithm, information theory, statistical mechanics, Hamiltonian dynamics

*Electronic address: dkuic@pmfst.hr
†Electronic address: pasko@pmfst.hr
‡Electronic address: juretic@pmfst.hr

I. INTRODUCTION

Maximum-entropy formalism, or alternatively MaxEnt algorithm, was formulated by E. T. Jaynes in his influential papers [1, 2] intended for applications in statistical mechanics. In Jaynes' approach a full development of the results of equilibrium statistical mechanics and formalism of Gibbs [3] was possible as a form of statistical inference based on the Shannon's concept of information-theory entropy as a measure of information [4]. In the language of Jaynes, it is the correct measure of the "amount of uncertainty" in the probability distribution [5]. Maximization of information-theory entropy subject to certain constraints is a central concept in Jaynes' approach and provides the least biased probability estimates subject to the available information. It is important that Jaynes sees Gibbs' formalism as essential tool for statistical inference in different problems with insufficient information, and this includes equilibrium statistical mechanics [1] and the formulation of a theory of irreversibility [2] that he tries to accomplish in his later works [5–9].

Predictions and calculations for different irreversible processes usually involve three distinct stages [6]: (1) Setting up an "ensemble", i.e., choosing an initial density matrix, or in our case an N-particle distribution function, which is to describe our initial knowledge about the system of interest; (2) Solving the dynamical problem; i.e., applying the microscopic equations of motion to obtain the time evolution of the system; (3) Extracting the final physical predictions from the time developed ensemble. As fully recognized by Jaynes, the stage (1) and the availability of its general solution simplifies the complicated stage (2). The problem includes also an equally important stage (0) consisting of some kind of measurement or observation defining both the system and the problem [10].

The aim of this paper is to consider the important stage (2) of the problem and assumptions that allow the use of MaxEnt inference methods, consistently with the Hamiltonian formulation of classical mechanics, for prediction of time evolution of closed macroscopic systems. In direct mathematical attempts leading to a theory of irreversibility the Liouville theorem with the conservation of phase space volume inherent to Hamiltonian dynamics is represented as one of the main difficulties. A necessary consequence of Hamiltonian dynamics and the Liouville theorem is that time evolution of the phase space probability density functions is governed by the Liouville equation. If, at the microscopic level Hamiltonian dynamics is invariant to time reversal, time evolution of phase space probability distributions that is determined from the Liouville equation of motion is reversible. For all these reasons alone, and the fact that one of the central mathematical problems in statistical mechanics is obtaining irreversible behavior for macroscopic systems from this immensely complicated equation of motion, it is reduced to an irreversible equation termed Boltzmann equation, rate equation or master equation. Far from creating difficulties, Jaynes considers the Liouville equation and the related constancy in time of Gibbs' entropy as precisely the dynamical property needed for solution of this problem considering it to be more of conceptual than mathematical nature [6].

Mathematical clarity of this viewpoint has its basis in a limit theorem noted by Shannon [4], an application of the fundamental asymptotic equipartition theorem of information theory, relating the Boltzmann's original formula for entropy of a macrostate and the Gibbs expression for entropy in the limit of a large number of particles [6, 7, 9]. Mathematical connection with the Boltzmann's interpretation of entropy as the logarithm of the number or ways (or microstates) by which a macroscopic state can be realized, $S = k \log W$, introduces then a simple physical interpretation to the Gibbs' formalism and its generalizations in the maximum-entropy formalism. Maximization of the information entropy subject to constraints predicts then the macroscopic behavior that can happen in the greatest number of ways compatible with the available information. In applications to time dependent processes this is referred to by Jaynes as the maximum caliber principle [8, 9]. Jaynes clearly stated that this does not represent a physical theory that explains the behavior of different systems by deductive reasoning from the first principles, but a form of statistical inference that makes predictions of observable phenomena from incomplete information [8]. For this reason predictive statistical mechanics can not claim deductive certainty for its predictions. This does not mean that it ignores the laws of microphysics; it certainly uses everything known about the structure of microstates and any data on macroscopic quantities, without making any extra physical assumptions beyond what is given by available information. It is important to note that sharp, definite predictions of macroscopic behavior are possible only when it is characteristic of each of the overwhelming majority of microstates compatible with data. For the same reason, this is just the behavior that is reproduced experimentally under those constraints; this is known essentially as the principle of macroscopic uniformity [1, 2], or reproducibility [9]. For predicting macroscopic behavior that is found to be reproducible it can not be necessary to specify all of the vast number of microvariables that were not under control and would not be the same on successive repetitions of the experiment [9]. In somewhat different context this property is recognized as the concept of macroscopic determinism whose precise definition involves some sort of thermodynamic limit [11]. In Jaynes' view the dynamical invariance of the Gibbs' entropy gives a simple proof of the second law, which is then a special case of a general requirement for any macroscopic process to be experimentally reproducible [6]. In the simple demonstration based on the Liouville theorem this makes possible for Jaynes to generalize the second law beyond the restrictions of initial and final equilibrium states, by considering it a special case of a general restriction on the direction of any reproducible process [6, 12]. The arguments used in this demonstration imply also

the question how to determine which nonequilibrium or equilibrium states can be reached from others, and this is not possible without information about dynamics, constants of motion, constraints, etc. The second law predicts only that a change of macroscopic state will go in the general direction of greater final entropy [9]. Better predictions are possible only by introducing more information. For predictions this is not hopeless; macrostates of higher entropy can be realized in overwhelmingly more ways and this is the reason for high reliability of the Gibbs equilibrium predictions [9]. In this context Jaynes also speculated that accidental success in reversal of an irreversible process is exponentially improbable [12].

Jaynes' interpretation of irreversibility and the second law reflects the point of view of the actual experimenter. Zurek [13] has introduced algorithmic randomness as the measure of the complexity of the microscopic state. He has prescribed entropy not only to the ensemble but also to the microscopic state. This prescription makes the principal distinction between his and Jaynes' approach. The basic laws of computation reflected in this interpretation allow Zurek to formulate thermodynamics from the point of view of Maxwell demon-type entities that can acquire information through measurements and process it in a manner analogous to Turing machines. According to Jaynes, the detailed description of microscopic development of the system can not be extracted from the data about macroscopic development, and therefore it is not a subject of his approach. Increase of entropy is related to gradual decrease of information about possible microstates of the system compatible with data.

MaxEnt algorithm and its standard methods represent a way of assigning probability distributions with the largest uncertainty and extent compatible with the available information, and for the same reasons, least biased with respect to unavailable information. Inferences drawn in this way depend only on our state of knowledge.[1] Macroscopic irreversibility is also an experimental fact and the nature of its subjectivity or objectivity is not necessarily a central issue when applying the MaxEnt algorithm in the spirit of Jaynes approach. In this work two different applications of MaxEnt algorithm to macroscopic closed systems with Hamiltonian dynamics and their time evolution are examined in detail along with their consequences. In the first approach Liouville equation is introduced as a strict *microscopic constraint* on time evolution of the phase space probability density functions, which are then completely determined by this constraint and their initial values. If probabilities are considered in the objective sense as a property of the system and not of our state of knowledge, this approach can be justified only if our knowledge of the microscopic dynamics is complete. As expected, this approach leads to the aforementioned well known problems, but represents a self-consistency check of the methods introduced in this variational problem to avoid problems with overconstraining. The concepts of *phase space paths*, *path probability distribution*, *conditional probability distribution*, and respective *path information entropy* and *conditional information entropy* are introduced. The importance of detailed dynamics of Hamiltonian systems and its representations by phase space paths is emphasized in maximization of the conditional information entropy subject to the constraint of Liouville equation for the conditional probability distribution. MaxEnt distribution and inferred time evolution then predicts macroscopic behavior than can happen in the greatest number of ways consistent with the information about microscopic dynamics, justified by our complete knowledge of it.

The concepts and quantities introduced in the first approach constitute also the basis for the second approach. The distinction is that Liouville equation for the conditional probability distribution is introduced as a *macroscopic constraint*. It means that it is regarded as a constraint on time evolution of the phase space probability density functions only on average, given by the integral over accessible phase space. In this respect it is similar to the constraints given by the data on macroscopic quantities. In Jaynes' predictive statistical mechanics more objectivity is ascribed to experimentally measured quantities than to probability distributions and their time evolution. The subjective aspect that becomes important here is that probabilities are assigned because of incomplete knowledge, i.e., partial information, and therefore represent our state of knowledge about the system. If our information about dynamics is not sufficient to determine the time evolution, an average is taken over all cases possible on the basis of our partial information. In this approach it can be seen how elements of irreversible macroscopic behavior in closed Hamiltonian systems appear as a consequence of gradual loss of information about possible microstates of the system. This idea has been developed by Jaynes in the density-matrix formalism [2]. Irreversible macroscopic behavior and the interpretation based on reproducibility and information loss has a clear mathematical description in the concepts of conditional information entropy and its relation with the information entropy, all introduced by Shannon [4]. The predicted final state of the system is recognized as the one characterized by maximum Boltzmann-Gibbs entropy. At the end of the work the subjective and objective aspects of the two approaches are indicated and relations with entropy production are discussed.

---

[1] This opinion is underlined with the following passage [2] "...if it can be shown that the class of phenomena predictable by maximum-entropy inference differs in any way from the class of experimentally reproducible phenomena, that fact would demonstrate the existence of new laws of physics, not presently known. Assuming that this occurs, and the new laws of physics are eventually worked out, and then the maximum-entropy inference based on the new laws will again have this property."

II. HAMILTONIAN DYNAMICS AND PHASE SPACE PATHS

The dynamical state of a Hamiltonian system with s degrees of freedom is described by the coordinates $q_1, q_2, \ldots, q_s$ and the momenta $p_1, p_2, \ldots, p_s$. At any time t it is represented by a point in the $2s$-dimensional Euclidean space Γ called the phase space of the system. To each possible state of the system corresponds a point in the phase space Γ, determined by the values of the coordinates $q_1, q_2, \ldots, q_s$ and the momenta $p_1, p_2, \ldots, p_s$. For a system of N point particles in three-dimensional space, s will be equal to $3N$. To simplify the notation, the symbol $(q, p) \equiv (q_1, \ldots, q_s, p_1, \ldots, p_s)$ is introduced for the set of $2s$ coordinates and momenta. The time dependence of $2s$ dynamical variables (q, p) is determined by Hamilton's equations

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \qquad 1 \leq i \leq s, \tag{1}$$

where $H \equiv H(q, p)$ is the Hamiltonian function equal to the energy of the system. Given the values (q_0, p_0) at some time t_0, Hamilton's equations (1) determine uniquely the values of the dynamical variables (q, p) at any other time t, given by the uniquely determined solution of equations (1),

$$q_i = q_i(t; q_0, p_0), \qquad p_i = p_i(t; q_0, p_0), \qquad 1 \leq i \leq s. \tag{2}$$

Motion of any point in the phase space Γ that represents the dynamical state of system, including all its subsequent or preceding time evolution, is thus determined. Any point (q, p) in the phase space Γ describes a curve called a *phase space path*, given by the uniquely determined solution of Hamilton's equations. Hence, through each point of the phase space Γ at any time t passes only one path. This is denoted by the index ω in $(q, p)_\omega$, where $\omega \in \Omega(\Gamma)$, and the set $\Omega(\Gamma)$ is the set of all paths in the phase space Γ. Index ω denotes this fact and is not used to parametrize the set $\Omega(\Gamma)$.

The velocity $v(q, p, t)$ of the point (q, p) in the phase space Γ at time t is given by

$$v \equiv |\mathbf{v}| = \sqrt{\sum_{i=1}^{s} \left(\frac{dq_i}{dt}\right)^2 + \sum_{i=1}^{s} \left(\frac{dp_i}{dt}\right)^2} = \sqrt{\sum_{i=1}^{s} \left(\frac{\partial H}{dp_i}\right)^2 + \sum_{i=1}^{s} \left(\frac{dH}{dq_i}\right)^2}. \tag{3}$$

The velocity vector $\mathbf{v}((q, p)_\omega, t)$ is tangential at the point $(q, p)_\omega \in \Gamma$ to the phase space path ω passing through it at time t. For the systems considered here the Hamiltonian function $H(q, p)$ does not depend on time and the velocity field $\mathbf{v}(q, p, t)$ of all points in the phase space Γ is stationary, i.e., $\mathbf{v}(q, p, t) \equiv \mathbf{v}(q, p)$.

Let M_0 be any measurable (in the sense of Lebesgue) set of points of the phase space Γ. In the Hamiltonian motion the set M_0 is transformed into another set M_t during an interval of time t. Liouville's theorem asserts that the measure of the set M_t for any t coincides with the measure of the set M_0.[2] This theorem proves that the measure in the phase space Γ,

$$\mu(M_t) = \int_{M_t} dq_1 \ldots dq_s dp_1 \ldots dp_s = \int_{M_t} d\Gamma, \tag{4}$$

is invariant under Hamiltonian motion. In the notation used in Eq. (4), volume element $dq_1 \ldots dq_s dp_1 \ldots dp_s$ of the phase space Γ is denoted by $d\Gamma$. One immediate consequence or corollary [14, pp. 18-19] of Liouville's theorem is that, if M_0 is a Lebesgue measurable set of points of the phase space Γ, of finite measure, and $f(q, p)$ a phase function Lebesgue integrable over Γ, then

$$\int_{M_t} f(q, p) d\Gamma = \int_{M_0} f(q(t; q_0, p_0), p(t; q_0, p_0)) d\Gamma_0. \tag{5}$$

Eq. (5) is obtained by changing the variables in the integral and introducing new variables (q_0, p_0) related to the variables (q, p) by transformation of the space Γ into itself in Hamiltonian motion, represented by uniquely determined solutions of Hamilton's equations in Eq. (2).

If, in particular, the set M is invariant to the Hamiltonian motion, then using this corollary it is easy to show how an integral of a phase function $f(q, p)$ over the invariant set M is transformed into an integration over the set $\Omega(M)$

[2] The theorem of Liouville is stated in this way in the book by Khinchin, [14, pp. 15-16].

of all paths in M. At any time t through each point $(q,p)_\omega \in \Gamma$ passes only one path $\omega \in \Omega(\Gamma)$. The path ω also passes through the point $(q_0,p_0)_\omega \in \Gamma$, given by the inverse of Eq. (2). The infinitesimal volume element $d\Gamma_0$ around the point $(q_0,p_0)_\omega$ can be written as $d\Gamma_0 = ds_{0\omega} dS_{0\omega}$. Here

$$(ds_{0\omega})^2 = \sum_{i=1}^{s} (dq_{0\omega i})^2 + \sum_{i=1}^{s} (dp_{0\omega i})^2, \tag{6}$$

is the (square of) infinitesimal distance along the path ω. The infinitesimal element $dS_{0\omega}$ of the surface $S_0(M)$ intersects the path ω perpendicularly at the point $(q_0,p_0)_\omega$. The surface $S_0(M)$ is then perpendicular to all paths in the set $\Omega(M)$ of all paths in M.[3]

The invariance of the measure $d\Gamma$ to Hamiltonian motion in Γ, and the fact that the velocity field $\mathbf{v}(q,p)$ in Γ is stationary as the Hamiltonian $H(q,p)$ does not depend on time, lead to the following consequence. *For any phase space path $\omega \in \Omega(\Gamma)$, the product of the velocity $v((q,p)_\omega)$ and the infinitesimal surface dS_ω intersecting the path ω perpendicularly at the point $(q,p)_\omega$, is constant under Hamiltonian motion along the entire length of the path ω*, i.e.,

$$v((q,p)_\omega) dS_\omega = const. \tag{7}$$

For any two points $(q_0,p_0)_\omega$ and $(q_a,p_a)_\omega$ on the same path ω, the following relation is obtained from (7):

$$v((q_0,p_0)_\omega) dS_{0\omega} = v((q_a,p_a)_\omega) dS_{a\omega}. \tag{8}$$

The infinitesimal element $dS_{a\omega}$ of the surface $S_a(M)$ intersects the path ω perpendicularly at the point $(q_a,p_a)_\omega$. Like the surface $S_0(M)$, surface $S_a(M)$ is also perpendicular to all paths in $\Omega(M)$. The infinitesimal elements $dS_{0\omega}$ and $dS_{a\omega}$ of the two surfaces $S_0(M)$ and $S_a(M)$ are connected by the path ω and neighboring paths determined by solutions of Hamilton's equations. The integral over surface $S_a(M)$ is transformed with the help of Eq. (8) into integration over surface $S_0(M)$,

$$\int_{S_a(M)} dS_{a\omega} = \int_{S_0(M)} \frac{v((q_0,p_0)_\omega)}{v((q_a,p_a)_\omega)} dS_{0\omega}. \tag{9}$$

Functional dependence between the points $(q_0,p_0)_\omega$ and $(q_a,p_a)_\omega$ on the path ω is not explicitly written in the integral (9); it is implied that this functional dependence is determined from solutions of Hamilton's equations.

Using Eq. (2) and keeping the times t and t_0 fixed allows the following notation to be introduced:

$$f(q(t;q_0,p_0),p(t;q_0,p_0)) \equiv g(q_0,p_0,t_0). \tag{10}$$

If notation in Eq. (10) (t and t_0 are fixed and the indices in (q_0,p_0) are replaced by the indices (q_a,p_a)) is substituted into the integral (5) over the set M which is invariant to Hamiltonian motion, this leads to following equality:

$$\int_M f(q,p) d\Gamma = \int_M g(q_a,p_a,t_0) d\Gamma_a = \int_M g(q_a,p_a,t_0) ds_{a\omega} dS_{a\omega}. \tag{11}$$

The integral (11) is then transformed using relation (8):

$$\int_M g(q_a,p_a,t_0) ds_{a\omega} dS_{a\omega} = \int_{S_0(M)} dS_{0\omega} v((q_0,p_0)_\omega) \int_\omega \frac{g(q_a,p_a,t_0)}{v(q_a,p_a)} ds_{a\omega}. \tag{12}$$

The function

$$F((q_0,p_0)_\omega,t_0) = v((q_0,p_0)_\omega) \int_\omega \frac{g(q_a,p_a,t_0)}{v(q_a,p_a)} ds_{a\omega}, \tag{13}$$

is defined on the surface $S_0(M)$ and is called a *path function* or *path distribution*. The integral in the relation (13) defining a path function $F((q_0,p_0)_\omega,t_0)$ is over the entire length of the path ω intersected perpendicularly by the

[3] It is a well established fact in the Hamiltonian formulation of classical mechanics that the integrals over subspaces of phase space of different dimensions form a sequence of canonical invariants known as the integral invariants of Poincaré [15, p. 403]. The volume integral (4) in the theorem of Liouville is a final member of this sequence. In this work, integrals over the surface $S_0(M)$ represent integration over the set of phase space paths $\Omega(M)$, and for that reason, a result given by Eq. (7) and a related noninvariant "measure" of the set $\Omega(M)$, Eq. (9), are important.

surface $S_0(M)$ at the point $(q_0, p_0)_\omega$. Infinitesimal element of the phase space path ω passing through the point $(q_a, p_a)_\omega$ is $ds_{a\omega}$, and the time t_0 in the integral (13) is fixed. There is no danger of the phase space velocity $v(q_a, p_a)$ in the integral (13) being equal to zero at some point along the path ω. If the phase space velocity $v(q_a, p_a)$ is equal to zero at some point $(q_a, p_a) \in \Gamma$, then the point (q_a, p_a) can not be part of any of phase space paths; points of this kind are called fixed points.

If the phase function $f(q, p)$ in Eq. (11) is a phase space probability density function, equal to zero everywhere outside the invariant set M, it is straightforward to prove that the path function $F((q_0, p_0)_\omega, t_0)$ defined by Eq. (13) satisfies the nonnegativity and normalization conditions required from probability distributions. Nonnegativity and normalization of the function $F((q_0, p_0)_\omega, t_0)$ which then represents a *path probability distribution* follow from the nonnegativity and normalization properties of the phase space probability density function $f(q, p)$. With the help of Eqs. (11) and (12) and the definition of $F((q_0, p_0)_\omega, t_0)$ in Eq. (13), one then obtains the normalization property

$$\int_M f(q,p)d\Gamma = \int_{S_0(M)} F((q_0,p_0)_\omega, t_0)dS_0 = 1. \tag{14}$$

Nonnegativity is established for all $(q_0, p_0)_\omega \in S_0(M)$ in a similar way. Integral over any invariant and measurable subset of the set M is transformed, in the way described above, into integral over a corresponding measurable subset on the surface $S_0(M)$. It is clear also that the measure defined on the surface $S_0(M)$ can be utilized as a measure on the set $\Omega(M)$ of all phase space paths in some invariant set M. Integrals in Eq. (14) are over all points $(q, p) \in M$, where $M \subset \Gamma$ is an invariant set, and equivalently, over all paths $\omega \in \Omega(M)$, passing through points $(q_0, p_0)_\omega \in S_0(M)$. Index ω denotes the fact that through each point $(q_0, p_0)_\omega \in S_0(M)$ at time t_0 passes only one path $\omega \in \Omega(M)$, since at any time through each point in Γ passes only one path.

III. MICROSTATE PROBABILITY AND PATH PROBABILITY

It is now possible to relate the microstate probability and the path probability in the phase space Γ of the system. Let the function $f(q, p, t)$ be a probability density function on Γ. The probability that at time t the system is in the element $d\Gamma$ of the $2s$-dimensional phase space Γ is equal $f(q, p, t)d\Gamma$. All points in the phase space Γ move according to Hamilton's equations (1) and the probability density function $f(q, p, t)$ satisfies the Liouville equation

$$\frac{\partial f}{\partial t} + \sum_{i=1}^{s}\left(\frac{\partial f}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial H}{\partial q_i}\right) \equiv \frac{df}{dt} = 0. \tag{15}$$

Since df/dt is a total or hydrodynamic derivative, Eq. (15) expresses that the time rate of change of $f(q, p, t)$ is zero along any phase space path given by uniquely determined solution of Hamilton's equations. In the notation used here, this fact is also written as

$$f((q,p)_\omega, t) = f((q_0,p_0)_\omega, t_0), \tag{16}$$

where points on the path $\omega \in \Omega(\Gamma)$ are related by Eq. (2) (here, specifically point $(q_0, p_0)_\omega$ at time t_0 and point $(q, p)_\omega$ at time t) .

In addition to the definition of the path probability distribution $F((q_0, p_0)_\omega, t_0)$ via microstate probability density function $f(q, p, t)$, it is possible to give another equivalent definition of $F((q_0, p_0)_\omega, t_0)$. In order to accomplish this, probability density function $\mathcal{F}(q, p, t; q_0, p_0, t_0)$ is introduced on the $4s$-dimensional Euclidean space $\Gamma \times \Gamma$. This function has the following special properties. If the integral of $\mathcal{F}(q, p, t; q_0, p_0, t_0)$ is taken over the phase space Γ with (q, p) as the integration variables, it gives the microstate probability density function $f(q_0, p_0, t_0)$ at time t_0,

$$\int_\Gamma \mathcal{F}(q,p,t;q_0,p_0,t_0)d\Gamma = f(q_0,p_0,t_0). \tag{17}$$

If $\mathcal{F}(q, p, t; q_0, p_0, t_0)$ is integrated over Γ with (q_0, p_0) as the integration variables this gives the microstate probability density function $f(q, p, t)$ at time t,

$$\int_\Gamma \mathcal{F}(q,p,t;q_0,p_0,t_0)d\Gamma_0 = f(q,p,t). \tag{18}$$

It is straightforward to prove that relations (17) and (18) are satisfied if the function $\mathcal{F}(q, p, t; q_0, p_0, t_0)$ has the following form:

$$\mathcal{F}(q,p,t;q_0,p_0,t_0) = f(q,p,t)\prod_{i=1}^{s}\delta(q_i - q_i(t;q_0,p_0))\delta(p_i - p_i(t;q_0,p_0)), \tag{19}$$

where $q_i(t;q_0,p_0)$ and $p_i(t;q_0,p_0)$ are the solutions (2) of Hamilton's equations and δ-s are Dirac delta functions. This is clear if one uses the relation (16) remembering that at times t and t_0 the solutions of Hamilton's equations $q_i(t;(q_0,p_0)_\omega)$ and $p_i(t;(q_0,p_0)_\omega)$ relate points $(q,p)_\omega$ and $(q_0,p_0)_\omega$ on the same phase space path $\omega \in \Omega(M)$. The function $\mathcal{F}(q,p,t;q_0,p_0,t_0)$ given by Eq. (19) satisfies the normalization condition

$$\int_\Gamma \int_\Gamma \mathcal{F}(q,p,t;q_0,p_0,t_0) d\Gamma d\Gamma_0 = 1. \tag{20}$$

In the space $\Gamma \times \Gamma$ function $\mathcal{F}(q,p,t;q_0,p_0,t_0)$ given by Eq. (19) represents the probability density that the point corresponding to the state of the system is in the element $d\Gamma_0$ around the point (q_0,p_0) at time t_0 and in the element $d\Gamma$ around the point (q,p) at time t.

As explained in Eq. (18), microstate probability density function $f(q,p,t)$ is given by the integral of the function $\mathcal{F}(q,p,t;q_0,p_0,t_0)$ over Γ with (q_0,p_0) as integration variables. Now, by applying the procedure similar to the one introduced in relations (11) and (12), the integral (18) can also be written as

$$f(q,p,t) = \int_{S_0(M)} dS_{0\omega} v((q_0,p_0)_\omega) \int_\omega \frac{\mathcal{F}(q,p,t;q_a,p_a,t_0)}{v(q_a,p_a)} ds_{a\omega}. \tag{21}$$

The infinitesimal distance $ds_{a\omega}$ is along the path ω and dS_0 is an infinitesimal element of $S_0(M)$, the surface which is perpendicular to all paths $\omega \in \Omega(M)$. The set M of all points in the phase space Γ that represent possible microstates of the system is invariant to Hamiltonian motion. The set $\Omega(M)$ of all phase space paths in M, determined by solutions (2) of Hamilton's equations, represents all possible time evolutions of the system.

Along with the lines leading to Eq. (21), function $G(q,p,t;(q_0,p_0)_\omega,t_0)$ is also introduced:

$$G(q,p,t;(q_0,p_0)_\omega,t_0) = v((q_0,p_0)_\omega) \int_\omega \frac{\mathcal{F}(q,p,t;q_a,p_a,t_0)}{v(q_a,p_a)} ds_{a\omega}. \tag{22}$$

Using the definition of $G(q,p,t;(q_0,p_0)_\omega,t_0)$ in Eq. (22), relation (21) is written as

$$f(q,p,t) = \int_{S_0(M)} G(q,p,t;(q_0,p_0)_\omega,t_0) dS_0. \tag{23}$$

The integral in the definition of $G(q,p,t;(q_0,p_0)_\omega,t_0)$ in Eq. (22) is over the entire length of the phase space path ω intersected perpendicularly by the surface $S_0(M)$ at the point $(q_0,p_0)_\omega$. It is clear that the expression

$$G(q,p,t;(q_0,p_0)_\omega,t_0) dS_0 d\Gamma \equiv \mathrm{P}(q,p,t \cap (q_0,p_0)_\omega,t_0), \tag{24}$$

represents the probability that the point corresponding to the state of the system is at time t_0 anywhere along the paths which pass through an infinitesimal element dS_0 around (q_0,p_0) on the surface $S_0(M)$ and that this point at some different time t is in the volume element $d\Gamma$ around (q,p).

Another definition of the path probability distribution $F((q_0,p_0)_\omega,t_0)$, in addition to Eq. (13), is possible in this way. It is given by the integral

$$F((q_0,p_0)_\omega,t_0) = \int_\Gamma G(q,p,t;(q_0,p_0)_\omega,t_0) d\Gamma. \tag{25}$$

The expression $F((q_0,p_0)_\omega,t_0) dS_0$ gives the probability that at time t_0 the point corresponding to the state of the system is anywhere along the paths that pass through an infinitesimal element dS_0 around the point $(q_0,p_0)_\omega$ on the surface $S_0(M)$. As before, measure on the set $\Omega(M)$ of all phase space paths in some invariant set M is defined as the measure on the surface $S_0(M)$. In accordance with the theory of probability, the ratio

$$\frac{G(q,p,t;(q_0,p_0)_\omega,t_0) dS_0 d\Gamma}{F((q_0,p_0)_\omega,t_0) dS_0} \equiv \mathrm{P}(q,p,t|(q_0,p_0)_\omega,t_0), \tag{26}$$

represents the *conditional probability* that at time t the point corresponding to the state of the system is in the element $d\Gamma$ around (q,p), if at time t_0 the point corresponding to the state of the system is anywhere along the paths passing through the infinitesimal element dS_0 around (q_0,p_0) on the surface $S_0(M)$. Relation (25) then proves that the integral of (26) over Γ with (q,p) as integration variables satisfies the normalization condition, i.e.

$$\int_\Gamma \frac{G(q,p,t;(q_0,p_0)_\omega,t_0) dS_0}{F((q_0,p_0)_\omega,t_0) dS_0} d\Gamma = 1. \tag{27}$$

To set up all the tools of probability theory needed in the rest of the work, conditional probability distribution $D(q,p,t|(q_0,p_0)_\omega,t_0)$ on the phase space Γ, which corresponds to the conditional probability (26), is defined by the relation

$$D(q,p,t|(q_0,p_0)_\omega,t_0)=\frac{G(q,p,t;(q_0,p_0)_\omega,t_0)}{F((q_0,p_0)_\omega,t_0)}. \tag{28}$$

Definition (28) of the conditional probability distribution $D(q,p,t|(q_0,p_0)_\omega,t_0)$ is possible and consistent only for paths passing through infinitesimal surface elements around the set of points $\{(q_0,p_0)\mid(q_0,p_0)\in S_0(M),F((q_0,p_0)_\omega,t_0)\neq 0\}$ on the surface $S_0(M)$ at time t_0. In the rest of the work, it is of practical importance to be able to write consistently

$$G(q,p,t;(q_0,p_0)_\omega,t_0)=D(q,p,t|(q_0,p_0)_\omega,t_0)F((q_0,p_0)_\omega,t_0). \tag{29}$$

For this reason, it is set here $D(q,p,t|(q_0,p_0)_\omega,t_0)=0$ for paths passing through infinitesimal surface elements around the remaining set of points $\{(q_0,p_0)\mid(q_0,p_0)\in S_0(M),\ F((q_0,p_0)_\omega,t_0)=0\}$ on the same surface $S_0(M)$.

Hamiltonian motion of all points in Γ is along the flow lines, i.e along the uniquely determined phase space paths $\omega\in\Omega(\Gamma)$. Liouville's theorem asserts that the measure of any measurable set of points of the phase space Γ is invariant under Hamiltonian motion. The relation (26), like the relation (24), represents probability and it is conserved in the phase space Γ. The total time derivative (i.e., time rate of change along the Hamiltonian flow lines) of this probability is equal to zero. In the relation (26) for the conditional probability, the path probability distribution $F((q_0,p_0)_\omega,t_0)$ and the surface element dS_0 are independent of the variables t and (q,p). Also, measure $d\Gamma$ is invariant to Hamiltonian motion. Therefore, it follows that the total time derivative of the conditional probability (26) is equal to zero if and only if

$$\frac{dG}{dt}\equiv\frac{\partial G}{\partial t}+\sum_{i=1}^{s}\left(\frac{\partial G}{\partial q_i}\frac{\partial H}{\partial p_i}-\frac{\partial G}{\partial p_i}\frac{\partial H}{\partial q_i}\right)=0\ . \tag{30}$$

This is a simple demonstration that the probability distribution $G(q,p,t;(q_0,p_0)_\omega,t_0)$ satisfies an equation analogous to the Liouville equation (15) for the microstate probability distribution $f(q,p,t)$. In relation to the macroscopic time evolution this issue is further discussed in this work.

IV. INFORMATION ENTROPIES AND MAXENT ALGORITHM

Different concepts introduced in previous sections are needed in finding relations between the Hamiltonian dynamics and the description of behavior of macroscopically large systems. Phase space paths given by uniquely determined solutions of Hamilton's equations represent information about the detailed microscopic dynamics of the Hamiltonian system. Microstate probability density functions on the other hand represent statistical, probabilistic aspect of description. The connection between two aspects is the Liouville equation governing the time evolution of microstate probability density functions, including all predictions on the values of macroscopic physical observables. In further sections that connection is studied with a maximum possible consistency, directed to prediction of the macroscopic time evolution of Hamiltonian closed systems based on the MaxEnt algorithm. Probability distribution utilized in this application of MaxEnt algorithm is the conditional probability distribution introduced in the previous section (Eqs. (28) and (29)). The reasoning behind this is that the conditional probability distribution $D(q,p,t|(q_0,p_0)_\omega,t_0)$ represents both described aspects, i.e., information about the microscopic dynamics represented by phase space paths and the probabilistic description which should be compatible with the macroscopic data. In general, MaxEnt algorithm is applied to ensure that the probability distribution describes only the information that we have and contains no hidden assumptions and information that is not available.

The quantity of the form $H=-\sum_i p_i\log p_i$ has a central role in information theory as a measure of information, choice and uncertainty for different probability distributions p_i. This quantity is named entropy of a probability distribution by Shannon [4] and is known also as information entropy. In an analogous manner Shannon defines entropy of a continuous distribution and entropy of N-dimensional continuous distribution. As pointed out by Jaynes [5], the analog of $-\sum_{i=1}^{n}p_i\log p_i$ for a discrete probability distribution p_i which goes over in a limit of infinite number of points $n\to\infty$ into a continuous distribution $w(x)$ (in such a way that the density of points divided by the total number of points approaches a definite function $m(x)$) is given by $-\int w(x)\log\left[\frac{w(x)}{m(x)}\right]dx$. Shannon [4] assumed the analog $-\int w(x)\log[w(x)]dx$, but he also pointed out an important difference between his definitions of the discrete and continuous entropies. The difference is that the former is absolute and the latter relative to the coordinate system. If we change coordinates the entropy of a continuous distribution will in general change in the way taken into

account by Shannon [4]. To achieve the required invariance of entropy of a continuous distribution under a change of the independent variable it is necessary to introduce a modification that follows from mathematical deduction. Accordingly, this is achieved with an introduction of the measure function $m(x)$ and yields the invariant information measure [5]:

$$S_I = -\int w(x) \log \left[\frac{w(x)}{m(x)}\right] dx. \tag{31}$$

In a limit going from the discrete distribution to a continuous distribution of several variables these results are readily generalized. If a uniform measure $m = const$ is assumed,[4] the invariant information measure differs from the Shannon's definition of entropy of a continuous distribution [4] by a irrelevant additive constant. Therefore, in all further discussions in this work, the irrelevant additive constant will be dropped.

Shannon [4] has also defined the joint and conditional entropies of a joint distribution of two continuous variables (which may themselves be multidimensional), concepts which are needed and utilized in this work. In the previous section, *joint distribution* $G(q,p,t;(q_0,p_0)_\omega,t_0)$ of two continuous multidimensional variables $(q,p) \in \Gamma$ and $(q_0,p_0)_\omega \in S_0(M)$ was introduced. Following the detailed explanation of Eq. (24), $G(q,p,t;(q_0,p_0)_\omega,t_0)d S_0 d\Gamma$ represents the probability of the joint occurrence of two events; the first occurring at time t_0 among the set of all possible phase space paths $\Omega(M)$ and the second occurring at time t among the set of all possible phase space points, the invariant set M. In accordance with Shannon's definition [4], *joint information entropy* of the joint distribution $G(q,p,t;(q_0,p_0)_\omega,t_0)$ is given by

$$S_I^G(t,t_0) = -\int_{S_0(M)} \int_\Gamma G \log G \; d\Gamma dS_0. \tag{32}$$

The notation $S_I^G(t,t_0)$ indicates that this quantity is a function of the time t and time t_0, through the distribution $G \equiv G(q,p,t;(q_0,p_0)_\omega,t_0)$. Following Shannon's definition [4], *conditional information entropy* of the joint distribution $G(q,p,t;(q_0,p_0)_\omega,t_0)$ is given by

$$S_I^{DF}(t,t_0) = -\int_{S_0(M)} \int_\Gamma G \log \left[\frac{G}{F}\right] \; d\Gamma dS_0, \tag{33}$$

where $F \equiv F((q_0,p_0)_\omega,t_0)$ is the path probability distribution. Using the definition of $D \equiv D(q,p,t|(q_0,p_0)_\omega,t_0)$ in Eqs. (28) and (29), one then immediately obtains the equivalent form of the conditional information entropy (33):

$$S_I^{DF}(t,t_0) = -\int_{S_0(M)} \int_\Gamma DF \log D \; d\Gamma dS_0. \tag{34}$$

From Eq. (34) it is clear that the conditional information entropy $S_I^{DF}(t,t_0)$ is the average of the entropy of conditional probability $D(q,p,t|(q_0,p_0)_\omega,t_0)$, weighted over all possible phase space paths $\omega \in \Omega(M)$ according to the path probability distribution $F((q_0,p_0)_\omega,t_0)$. More explicitly, it is the average of the entropy of $(q,p) \in \Gamma$ at time t for each value of $(q_0,p_0) \in S_0(M)$, weighted according to the probability distribution $F((q_0,p_0)_\omega,t_0)$.

Relation between the information entropies $S_I^G(t,t_0)$ and $S_I^{DF}(t,t_0)$, introduced in Eqs. (32) and (33), is completed by introducing the information entropy of the distribution $F((q_0,p_0)_\omega,t_0)$, or alternatively, *path information entropy*:

$$S_I^F(t_0) = -\int_{S_0(M)} F \log F \; dS_0. \tag{35}$$

Relation between $S_I^G(t,t_0)$, $S_I^{DF}(t,t_0)$ and $S_I^F(t_0)$ is obtained straightforwardly, if one substitutes Eq. (29) in Eq. (32) for the information entropy $S_I^G(t,t_0)$, and then uses the properties of probability distributions. In this way one obtains

$$S_I^G(t,t_0) = S_I^{DF}(t,t_0) + S_I^F(t_0). \tag{36}$$

Relation (36) between $S_I^G(t,t_0)$, $S_I^{DF}(t,t_0)$ and $S_I^F(t_0)$, in accordance with an analogous relation of Shannon [4], states that the uncertainty (or entropy) of the joint event is equal to the uncertainty of the first plus the uncertainty

---

[4] In classical statistical mechanics one assumes a uniform measure for phase space: $m = const$. The measure for phase space is of considerable interest as we can regard classical statistical mechanics as a limiting form of quantum statistical mechanics (for further reference see [5]).

of the second when the first is known (and vice versa). This is the interpretation of relation (36), with the first event occurring at time t_0 among the set of all possible phase space paths $\Omega(M)$, and the second event occurring at time t among the set of all possible phase space points, the invariant set M.

It is important to give additional comments to Eq. (36). In general, uncertainty of the joint event is less then or equal to the sum of uncertainties of the two individual events, with the equality if (and only if) the two events are independent [4]. The probability distribution of the joint event is given here by $G(q,p,t;(q_0,p_0)_\omega,t_0)$. Information entropy or uncertainty of one of them (in this case called the second event because of its occurrence at a later time) is given by

$$S_I^f(t) = -\int_\Gamma f \log f \; d\Gamma. \tag{37}$$

The quantity $S_I^f(t)$ is also known as the information entropy of the microstate probability distribution $f(q,p,t)$, or in short, *information entropy*. The uncertainty of the first event is the path information entropy $S_I^F(t_0)$, defined in Eq. (35). The aforementioned property of information entropies is given here for $S_I^G(t,t_0)$, $S_I^f(t)$ and $S_I^F(t_0)$ in the following relation:

$$S_I^G(t,t_0) \leq S_I^f(t) + S_I^F(t_0), \tag{38}$$

with the equality if (and only if) the two events are independent. Furthermore, from Eqs. (36) and (38) one obtains an important relation between $S_I^f(t)$ and $S_I^{DF}(t,t_0)$:

$$S_I^f(t) \geq S_I^{DF}(t,t_0), \tag{39}$$

with the equality if (and only if) the events occurring at time t_0 among the set of all possible phase space paths $\Omega(M)$ and at time t among the set of all possible phase space points M, are independent of each other.

Time evolution of the joint probability distribution $G(q,p,t;(q_0,p_0)_\omega,t_0)$ is determined by the Hamiltonian motion via Eq. (30) which is analogous to the Liouville equation (15) for the microstate probability distribution $f(q,p,t)$. This is explicit in another way if one considers the definition of $G(q,p,t;(q_0,p_0)_\omega,t_0)$ in Eq. (22), where the information about Hamiltonian dynamics enters via Eqs. (16) and (19). Possible microstates in Γ at time t are dependent on the possible phase space paths at time t_0, and vice versa, as through any point of Γ at any time t passes only one path. If the values of joint probability distribution $G(q,p,t;(q_0,p_0)_\omega,t_0)$ are well defined for all points $(q,p) \in \Gamma$ and $(q_0,p_0) \in S_0(M)$ at given initial time $t = t_0$, its values are then determined at all times t in the entire phase space Γ via Liouville equation (30). In the terms of probability, the events occurring at time t_0 among the set of all possible phase space paths $\Omega(M)$ and at time t among the set of all possible phase space points $M \subset \Gamma$, are not independent. Deduction based on these arguments, suggests that maximization of the conditional information entropy $S_I^{DF}(t,t_0)$ subject to the constraints of Liouville Eq. (30) and normalization, can not yield the upper bound which is given (at any time t) by the value of the information entropy $S_I^f(t)$. Attaining this upper bound (i.e., $S_I^{DF}(t,t_0) = S_I^f(t)$) would require statistical independence or, in other words, a complete loss of correlation between the set of possible phase space paths $\Omega(M)$ at time t_0 and the set of possible phase space points $M \subset \Gamma$ at time t. This is precluded at any time t by the constraint implied by Liouville Eq. (30).

At this point it is helpful to make a distinction between two aspects of time evolution. The first is a microscopic aspect which represents a problem of dynamics implied in this work by Hamilton's equations. The solutions are represented in Γ as phase space paths. Predicting a macroscopic time evolution represents a problem of available information and drawing inferences from that partial information. To be precise, in the problem considered here, this is finding the set of possible phase space paths $\Omega(M)$ at time t_0 and relating them to the set of possible phase space points $M \subset \Gamma$ at time t, along with the probability of their occurrence. Therefore, the microscopic dynamics and the respective phase space paths are also part of information problem. Information about microscopic dynamics in a case of macroscopic system is very likely to be incomplete for a variety of different possible reasons. However, in the absence of more complete knowledge, Hamilton's equations (1) and the set of possible phase space paths are the representation of our information about microscopic dynamics. It is natural to assume that the predicted macroscopic time evolution for a closed system is consistent with our knowledge of microscopic dynamics, even if this knowledge is not complete. In accordance with this the following assumptions are made: *all phase space paths in the set M of possible points in Γ are given entirely by the set $\Omega(M)$ of all possible phase space paths at time t_0. The set M of all possible phase space points is an invariant set. The set $\Omega(M)$ is comprised of all possible phase space paths at any time $t \geq t_0$ and is invariant to time evolution.*

All the reasons mentioned before lead to a conclusion that regarding Liouville Eq. (30) as a strict *microscopic constraint* on time evolution is equivalent to having complete knowledge of microscopic dynamics. Following previously introduced assumptions, Liouville Eq. (30) can also be regarded in a different way as a *macroscopic constraint* on time

evolution. If our information about microscopic dynamics is not sufficient to determine the time evolution, an average is taken over all cases possible on the basis of our partial information. The conditional information entropy $S_I^{DF}(t,t_0)$ is maximized subject to the constraint of Liouville Eq. (30) regarded as a phase space average, or more precisely, an integral over phase space similarly to other macroscopic constraints. In this way "objectivity" of probability assignment for the conditional probability distribution $D(q,p,t|(q_0,p_0)_\omega,t_0)$ and predictions made is ensured from introducing extraneous assumptions not warranted by our data. In predictive statistical mechanics as formulated by Jaynes, inferences are drawn from probability distributions whose sample spaces represent what is known about the structure of microstates, and maximize information entropy subject to the available macroscopic data [8]. By developing this description further, loss of correlation between the initial phase space paths and final microstates leads to uncertainty in prediction, which in this sense becomes a property of MaxEnt inferences on macroscopic time evolution. Measure of this uncertainty is given by the conditional information entropy $S_I^{DF}(t,t_0)$.

V. MAXENT INFERENCES AND TIME EVOLUTION

It is assumed that it is possible to assign the path probability distribution $F((q_0,p_0)_\omega,t_0)$ at initial time t_0 to the set $\Omega(M)$ of possible paths in the phase space Γ. The initial path probability distribution distribution $F((q_0,p_0)_\omega,t_0)$ is compatible with the available information at time t_0. It may be assigned by MaxEnt algorithm applied on the path information entropy $S_I^F(t_0)$ given in Eq. (35), or some other means. Microstate probability distribution $f(q,p,t_0)$ at initial time t_0 is not completely determined from $F((q_0,p_0)_\omega,t_0)$; values of $f(q,p,t_0)$ for possible microstates are not completely assigned in this way. It is possible to calculate the values of $f(q,p,t)$ in Γ at any time $t \geq t_0$ from $F((q_0,p_0)_\omega,t_0)$ using Eqs. (23) and (29), but for that full knowledge of the conditional probability distribution $D(q,p,t|(q_0,p_0)_\omega,t_0)$ is necessary. This is due to the fact that phase space paths are one-dimensional objects in Γ, each comprising an infinite number of points and therefore permitting an infinite number of different initial microstates along each path. In other words, points representing possible states of the system at initial time are not precisely determined by the phase space paths although microscopic dynamics is; this requires more information.

In the first approach, time evolution of the conditional probability distribution $D(q,p,t|(q_0,p_0)_\omega,t_0)$ in the interval $t_0 \leq t \leq t_a$ should satisfy the following constraints: normalization condition

$$\int_M D(q,p,t|(q_0,p_0)_\omega,t_0)d\Gamma = 1. \tag{40}$$

and the Liouville equation for $D(q,p,t|(q_0,p_0)_\omega,t_0)$,

$$\frac{\partial D}{\partial t} + \sum_{i=1}^{s}\left(\frac{\partial D}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial D}{\partial p_i}\frac{\partial H}{\partial q_i}\right) = 0. \tag{41}$$

By Eq. (29), constraints given by the Liouville Eqs. (30) and (41) are equivalent. By definition, the set $M \subset \Gamma$ of all possible microstates is an invariant set. Normalization condition in Eq. (40) contains information about the structure of possible microstates in the phase space Γ, in the time interval under consideration $t_0 \leq t \leq t_a$. Information about microscopic dynamics is contained in the Liouville Eq. (41). Information which is available at initial time t_0 is contained in the path probability distribution distribution $F((q_0,p_0)_\omega,t_0)$, defined on the surface $S_0(M)$ perpendicular to all paths in the set $\Omega(M)$ of possible phase space paths.

Time derivative of the conditional information entropy $S_I^{DF}(t,t_0)$ is given by

$$\frac{dS_I^{DF}(t,t_0)}{dt} = -\int_{S_0(M)}\int_M \frac{\partial D}{\partial t}F\log D\ d\Gamma dS_0 - \int_{S_0(M)}\int_M \frac{\partial D}{\partial t}F\ d\Gamma dS_0. \tag{42}$$

Because of the normalization, Eq. (40), last term in Eq. (42) is equal to zero. As it will become clear, it is safe to use somewhat simplified expression with this term left out from the variational problem. At time t_a, conditional information entropy $S_I^{DF}(t_a,t_0)$ is then given by the expression

$$S_I^{DF}(t_a,t_0) = -\int_{t_0}^{t_a}\int_{S_0(M)}\int_M \frac{\partial D}{\partial t}F\log D\ d\Gamma dS_0 dt + S_I^{DF}(t_0,t_0). \tag{43}$$

Path probability distribution $F((q_0,p_0)_\omega,t_0)$ is fixed. The conditional probability distribution $D(q,p,t|(q_0,p_0)_\omega,t_0)$ is determined by the Liouville Eq. (41) and its initial values at time t_0. Therefore, $S_I^{DF}(t_0,t_0)$ in Eq. (43) is also varied in the maximization of $S_I^{DF}(t_a,t_0)$ at time t_a, subject to the constraints on time evolution of $D(q,p,t|(q_0,p_0)_\omega,t_0)$

given by the Liouville Eq. (41) and normalization (40). Restriction of the class of functions that render $S_I^{DF}(t_a,t_0)$ in Eq. (43) stationary subject to constrains (40) and (41), only to those functions with physically well defined values in Γ at initial time t_0 represents an intricate mathematical task. For that reason this problem is avoided here by forming the functional

$$J[D]=S_I^{DF}(t_a,t_0)-S_I^{DF}(t_0,t_0)=\int_{t_0}^{t_a}\int_{S_0(M)}\int_M K(D,\partial_t D)d\Gamma dS_0 dt, \tag{44}$$

with the function $K(D,\partial_t D)$ given by

$$K(D,\partial_t D)=-\frac{\partial D}{\partial t}F\log D. \tag{45}$$

From the simple form of $K(D,\partial_t D)$ it follows, under assumption that the function $D(q,p,t|(q_0,p_0)_\omega,t_0)$ has continuous first partial derivatives, that the Euler equation with $K(D,\partial_t D)$ given by Eq. (45) is satisfied for all such functions,

$$\frac{dK}{dD}-\frac{d}{dt}\left(\frac{dK}{d(\partial_t D)}\right)-\sum_{i=1}^{s}\left[\frac{d}{dq_i}\left(\frac{dK}{d(\partial_{q_i}D)}\right)+\frac{d}{dp_i}\left(\frac{dK}{d(\partial_{p_i}D)}\right)\right]=0. \tag{46}$$

Eq. (46) is a simple mathematical identity which follows from the form of $K(D,\partial_t D)$, but the time evolution of $D(q,p,t|(q_0,p_0)_\omega,t_0)$ is determined in the constrained variational problem that is considered in this work. In the variational problem that is considered here, functional $J[D]$ in Eq. (44) should be stationary with respect to variations subject to the constraints (40) and (41). In addition, on the boundary of integration region $M\times(t_0,t_a)$ in (44), values of $D(q,p,t|(q_0,p_0)_\omega,t_0)$ are not prescribed. The constraints given by Eqs. (40) and (41) are valid for each point $(q_0,p_0)_\omega\in S_0(M)$, i.e., for each path $\omega\in\Omega(M)$, individually. They are written here in equivalent but more suitable form:

$$\varphi_1((q_0,p_0)_\omega,t_0;t,D)=F((q_0,p_0)_\omega,t_0)\int_M D(q,p,t|(q_0,p_0)_\omega,t_0)d\Gamma-F((q_0,p_0)_\omega,t_0)=0, \tag{47}$$

and

$$\varphi_2((q_0,p_0)_\omega,t_0;q,p,t,\partial_q D,\partial_p D,\partial_t D)=\left[\frac{\partial D}{\partial t}+\sum_{i=1}^{s}\left(\frac{\partial D}{\partial q_i}\frac{\partial H}{\partial p_i}-\frac{\partial D}{\partial p_i}\frac{\partial H}{\partial q_i}\right)\right]F=0. \tag{48}$$

Conditional probability distribution $D(q,p,t|(q_0,p_0)_\omega,t_0)$ that renders $J[D]$ stationary, subject to the isoperimetric (47) and strict equality constraints (48), is completely determined by the initial conditional probability distribution $D(q,p,t_0|(q_0,p_0)_\omega,t_0)$ at time t_0, and it is assumed here only that it is well defined in the sense of Eqs.(26) and (28).

Methods for variational problems with this type of constraints exist and one can develop them and apply in practical problems [16]. Here, in the notation which is adapted to this particular problem, the following functionals are introduced:

$$C_1[D,\lambda_1]=\int_{S_0(M)}\int_{t_0}^{t_a}\lambda_1\varphi_1\ dtdS_0, \tag{49}$$

and

$$C_2[D,\lambda_2]=\int_{S_0(M)}\int_{t_0}^{t_a}\int_M\lambda_2\varphi_2\ d\Gamma dtdS_0. \tag{50}$$

The Lagrange multipliers $\lambda_1\equiv\lambda_1((q_0,p_0)_\omega,t_0;t)$ and $\lambda_2\equiv\lambda_2((q_0,p_0)_\omega,t_0;q,p,t)$ are functions defined in the integration regions in (49) and (50). For any function with continuous first partial derivatives, Euler equation for the constraint $\varphi_2\equiv\varphi_2((q_0,p_0)_\omega,t_0;q,p,t,\partial_q D,\partial_p D,\partial_t D)$ is equal to zero. Following the multiplier rule for such problems explained in reference [16], we introduce an additional (constant) Lagrange multiplier λ_0 as a multiplicative factor for K,

$$J[D,\lambda_0]=\int_{t_0}^{t_a}\int_{S_0(M)}\int_M\lambda_0 K(D,\partial_t D)\ d\Gamma dS_0 dt. \tag{51}$$

The functional $I[D,\lambda_0,\lambda_1,\lambda_2]$ is formed from $J[D,\lambda_0]$, $C_1[D,\lambda_1]$ and $C_2[D,\lambda_2]$:

$$I[D,\lambda_0,\lambda_1,\lambda_2]=J[D,\lambda_0]-C_1[D,\lambda_1]-C_2[D,\lambda_2]. \tag{52}$$

Existence of Lagrange multipliers $\lambda_0 \neq 0$, λ_1 and λ_2, such that the variation of $I[D, \lambda_0, \lambda_1, \lambda_2]$ is stationary,

$$\delta I = \delta J - \delta C_1 - \delta C_2 = 0, \tag{53}$$

represents a proof that it is possible to make $J[D]$ in Eq. (44) stationary subject to constraints (47) and (48). Due to constraint (48), comparison functions considered in this variational problem should have at least their first partial derivatives continuous. The function $D(q, p, t|(q_0, p_0)_\omega, t_0)$ which renders $J[D]$ stationary subject to (47) and (48) must satisfy the Euler equation:

$$\lambda_0 \left\{ \frac{dK}{dD} - \frac{d}{dt}\left(\frac{dK}{d(\partial_t D)}\right) - \sum_{i=1}^{s}\left[\frac{d}{dq_i}\left(\frac{dK}{d(\partial_{q_i} D)}\right) + \frac{d}{dp_i}\left(\frac{dK}{d(\partial_{p_i} D)}\right)\right]\right\}$$
$$-\lambda_1 F + \left[\frac{\partial \lambda_2}{\partial t} + \sum_{i=1}^{s}\left(\frac{\partial \lambda_2}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial \lambda_2}{\partial p_i}\frac{\partial H}{\partial q_i}\right)\right] F = 0. \tag{54}$$

Because of Eq. (46), the term multiplied by λ_0 in Euler equation (54) is equal to zero. Stationarity of the functional $I[D, \lambda_0, \lambda_1, \lambda_2]$ in Eq. (52) is therefore possible even with $\lambda_0 \neq 0$. Then, following the multiplier rule [16], it is possible to set $\lambda_0 = 1$ in Eqs. (52), (53) and (54) without affecting the solution. This is equivalent to dividing by the constant Lagrange multiplier λ_0.

Taking into account Eq. (46), from Eq. (54) follows that the Lagrange multipliers λ_1 and λ_2 satisfy the equation

$$\left[\frac{\partial \lambda_2}{\partial t} + \sum_{i=1}^{s}\left(\frac{\partial \lambda_2}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial \lambda_2}{\partial p_i}\frac{\partial H}{\partial q_i}\right)\right] F = \lambda_1 F. \tag{55}$$

For phase space paths passing through infinitesimal surface elements around the set of points $\{(q_0, p_0) \mid (q_0, p_0) \in S_0(M),\ F((q_0, p_0)_\omega, t_0) \neq 0\}$ on the surface $S_0(M)$, Eq. (55) is equivalent to

$$\frac{\partial \lambda_2}{\partial t} + \sum_{i=1}^{s}\left(\frac{\partial \lambda_2}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial \lambda_2}{\partial p_i}\frac{\partial H}{\partial q_i}\right) = \lambda_1. \tag{56}$$

In accordance with the explanation of this variational problem given after Eq. (46), functions $D(q, p, t|(q_0, p_0)_\omega, t_0)$ that render $J[D]$ in Eq. (44) stationary subject to the constraints (47) and (48), are not required to take prescribed values on the boundary of integration region $M \times (t_0, t_a)$. Then, it is necessary that, in addition to satisfying the Euler equation (54), they also satisfy the Euler boundary condition on the boundary of integration region [16]. For all points on the part of the boundary of $M \times (t_0, t_a)$ where $t = t_0$ and $t = t_a$, Euler boundary condition gives independently:

$$\left[\frac{dK}{d(\partial_t D)} - \lambda_2 F\right]_{t=t_0,t_a} = -\left[\log D + \lambda_2\right]_{t=t_0,t_a} F = 0. \tag{57}$$

For all points on the part of the boundary of $M \times (t_0, t_a)$ where time $t_0 \leq t \leq t_a$ is allowed to vary, the Euler boundary condition gives:

$$F\left[\lambda_2 \mathbf{v} \cdot \mathbf{n}\right]_{\text{ at the boundary of } M} = 0. \tag{58}$$

In Eq. (58), $\mathbf{v} \cdot \mathbf{n}$ is a scalar product of the velocity field $\mathbf{v}(q, p)$ in Γ (defined in Sec. II) and the unit normal $\mathbf{n}$ of the boundary surface of invariant set M, taken at the surface. Eq. (58) is satisfied naturally due to Hamiltonian motion, since the set M is invariant by definition, and therefore $\mathbf{v} \cdot \mathbf{n} = 0$ for all points on the boundary surface of M. This is a consequence of the fact that phase space paths do not cross over the boundary surface of the invariant set M.

Eq. (46) is satisfied for all functions with continuous first partial derivatives, and the scalar product in Eq. (58) vanishes naturally due to Hamiltonian motion. Therefore, functions $D(q, p, t|(q_0, p_0)_\omega, t_0)$ that render $J[D]$ in Eq. (44) stationary subject to the constraints (47) and (48) are determined from these constraints and the boundary condition (57). From this boundary condition one obtains the form of $D(q, p, t|(q_0, p_0)_\omega, t_0)$ at times t_0 and t_a,

$$D(q, p, t|(q_0, p_0)_\omega, t_0)|_{t=t_0,t_a} = \exp\left[-\lambda_2((q_0, p_0)_\omega, t_0; q, p, t)\right]|_{t=t_0,t_a}. \tag{59}$$

Since it is only required that $t_a \geq t_0$, time t_a is arbitrary in other respects. The boundary condition (57) then holds for any time $t \geq t_0$:

$$D(q, p, t|(q_0, p_0)_\omega, t_0) = \exp\left[-\lambda_2((q_0, p_0)_\omega, t_0; q, p, t)\right]. \tag{60}$$

From the constraint (48), using Eq. (60), one immediately obtains an equation for the Lagrange multiplier $\lambda_2((q_0,p_0)_\omega,t_0;q,p,t)$:

$$\frac{\partial\lambda_2}{\partial t}+\sum_{i=1}^{s}\left(\frac{\partial\lambda_2}{\partial q_i}\frac{\partial H}{\partial p_i}-\frac{\partial\lambda_2}{\partial p_i}\frac{\partial H}{\partial q_i}\right)=0, \tag{61}$$

By comparison of Eq. (56) with Eq. (61), it follows that for all $t\geq t_0$,

$$\lambda_1((q_0,p_0)_\omega,t_0;t)=0. \tag{62}$$

Eqs. (59-62) refer to phase space paths passing through infinitesimal surface elements around the set of points $\{(q_0,p_0)\mid(q_0,p_0)\in S_0(M),\ F((q_0,p_0)_\omega,t_0)\neq 0\}$ on the surface $S_0(M)$. This issue is explained more fully in Section III, Eqs. (28) and (29).

The conditional probability distribution in Eq. (60) renders the functional $J[D]$ in Eq. (44) stationary subject to the constraints (47) and (48). With the help of the constraint (48) and Eqs. (60) and (61), it is then straightforward to check that the value of the conditional information entropy $S_I^{DF}(t,t_0)$ is constant for all times $t\geq t_0$. The value of $S_I^{DF}(t,t_0)$ in this way depends on the initial conditional probability distribution $D(q,p,t_0|(q_0,p_0)_\omega,t_0)$ at time t_0, which is only assumed to be physically well defined. Furthermore, for all conditional probability distributions that are well defined, the upper bound on $S_I^{DF}(t,t_0)$, given by Eq. (39), is not attained in maximization. This is not possible because it requires statistical independence, i.e., complete loss of correlation between the initial phase space paths and final microstates. This is precluded by the constraint (48) on the microscopic dynamics. Therefore, it is now possible to state that *complete knowledge of microscopic dynamics, implied in the constraint of Liouville equation (41), precludes the conditional information entropy $S_I^{DF}(t,t_0)$ from attaining the upper bound (39), which requires statistical independence.* The problem of information about microscopic dynamics, regarding all types of possible interactions, translates into a general question on objectivity or subjectivity of our probability assignments and inferences on macroscopic time evolution. Incomplete knowledge of microscopic dynamics leads to a loss of information about possible microstates of the system during reproducible time evolution compatible with our macroscopic data. The role of constraints is central in this respect, as it is clear in the second approach.

The above conclusions reached from the interpretation of relation (39) and the property of $S_I^{DF}(t,t_0)$ being a measure of uncertainty, as explained at the end of Section IV, form a basis for the second approach. Loss of correlation between the initial phase space paths and final microstates enters into description by regarding Liouville equation (41) for the conditional probability distribution $D(q,p,t|(q_0,p_0)_\omega,t_0)$ as a macroscopic constraint, along with other macroscopic constraints in Jaynes' theory. This is done by replacing the strict equality constraint (48) by the constraint which is of isoperimetric form,

$$\varphi_2((q_0,p_0)_\omega,t_0;t,D)=\int_M\left[\frac{\partial D}{\partial t}+\sum_{i=1}^{s}\left(\frac{\partial D}{\partial q_i}\frac{\partial H}{\partial p_i}-\frac{\partial D}{\partial p_i}\frac{\partial H}{\partial q_i}\right)\right]F\ d\Gamma=0. \tag{63}$$

The functional (50) is then replaced with the functional

$$C_2[D,\lambda_2]=\int_{S_0(M)}\int_{t_0}^{t_a}\lambda_2\varphi_2\ dtdS_0. \tag{64}$$

Lagrange multiplier $\lambda_2\equiv\lambda_2((q_0,p_0)_\omega,t_0;t)$ is now a function defined in the integration region in the functional (64). With replacement of the constraint (48) by the constraint (63), one can proceed with this variational problem in analogy with the first approach. Information that the set M of possible microstates is invariant to Hamiltonian motion is contained in the constraint (63). Analogy is not complete, since the constraint (63) is more flexible and does not determine the time evolution of $D(q,p,t|(q_0,p_0)_\omega,t_0)$ in the way that the constraint (48) does. In simple words, there is a much larger class of functions that satisfy the constraint (63), including all functions that, in addition, satisfy also the constraint (48). This broader class of functions allows for maximization of the conditional information entropy $S_I^{DF}(t_a,t_0)$ in Eq. (43) subject to constraints (47) and (63), even if the initial $D(q,p,t_0|(q_0,p_0)_\omega,t_0)$ at time t_0 is prescribed. In this way the initial conditional information entropy term $S_I^{DF}(t_0,t_0)$ in Eq. (43) is equal to some fixed value and, therefore, is not varied in this redefined variational problem. The prescribed initial $D(q,p,t_0|(q_0,p_0)_\omega,t_0)$ at time t_0 must be physically well defined.

In all other respects, one can proceed in complete analogy with the first approach. Due to constraint (63) comparison functions considered in this variational problem should have their first partial derivatives continuous. The function $D(q,p,t|(q_0,p_0)_\omega,t_0)$ that maximizes $S_I^{DF}(t_a,t_0)$ subject to constraints (47) and (63) satisfies the Euler equation:

$$\lambda_0\left\{\frac{dK}{dD}-\frac{d}{dt}\left(\frac{dK}{d(\partial_t D)}\right)-\sum_{i=1}^{s}\left[\frac{d}{dq_i}\left(\frac{dK}{d(\partial_{q_i}D)}\right)+\frac{d}{dp_i}\left(\frac{dK}{d(\partial_{p_i}D)}\right)\right]\right\}-\lambda_1F+\frac{\partial\lambda_2}{\partial t}F=0. \tag{65}$$

Eq. (46) still holds here. Therefore, in a completely analogous manner already explained in first approach, it is possible to set here $\lambda_0 = 1$ in Eq. (65) without affecting the solution. Eqs. (46) and (65) then lead to the equation for Lagrange multipliers $\lambda_1((q_0, p_0)_\omega, t_0; t)$ and $\lambda_2((q_0, p_0)_\omega, t_0; t)$, analogous to Eq. (56):

$$\frac{\partial \lambda_2}{\partial t} = \lambda_1. \tag{66}$$

Function $D(q, p, t|(q_0, p_0)_\omega, t_0)$ that maximizes $S_I^{DF}(t_a, t_0)$ subject to constraints (47) and (63) is required also to take prescribed values on the part of the boundary of integration region $M \times (t_0, t_a)$ where $t = t_0$. This function is not prescribed on the remaining part of the boundary of $M \times (t_0, t_a)$, and it is necessary that it satisfies the Euler boundary condition on that part of the boundary, [16]. For all points on the part of the boundary of $M \times (t_0, t_a)$ where $t = t_a$, the Euler boundary condition gives:

$$\left[\frac{dK}{d(\partial_t D)} - \lambda_2 F\right]_{t=t_a} = -\left[\log D + \lambda_2\right]_{t=t_a} F = 0. \tag{67}$$

The Euler boundary condition is satisfied naturally for all points on the part of the boundary of $M \times (t_0, t_a)$ where time $t_0 \leq t \leq t_a$ is allowed to vary. The set M is invariant to Hamiltonian motion and, in the way already explained in Eq. (58), analogous equation here is also satisfied naturally due to Hamiltonian motion.

The form of MaxEnt conditional probability distribution at time t_a follows from the boundary condition (67):

$$D(q, p, t_a|(q_0, p_0)_\omega, t_0) = \exp\left[-\lambda_2((q_0, p_0)_\omega, t_0; t_a)\right]. \tag{68}$$

In this way, MaxEnt conditional probability distribution $D(q, p, t|(q_0, p_0)_\omega, t_0)$ is not strictly determined in the interval $t_0 < t < t_a$, with the exception of times t_0 and t_a. This is a consequence of the fact that Eq. (46) is satisfied for all functions with continuous first partial derivatives and that the macroscopic constraint (63) does not determine time evolution of $D(q, p, t|(q_0, p_0)_\omega, t_0)$ in the way that the strict microscopic constraint (48) does. Therefore, *for any well defined initial conditional probability distribution $D(q, p, t_0|(q_0, p_0)_\omega, t_0)$ at time t_0, there is an entire class of equally probable solutions obtained by MaxEnt algorithm, all satisfying the macroscopic constraint (63). In the interval $t_0 < t < t_a$, all functions in this class describe possible time evolutions of the conditional probability distribution, compatible with the macroscopic constraint (63) and the initial conditional probability distribution $D(q, p, t_0|(q_0, p_0)_\omega, t_0)$ at time t_0. At time t_a, the form of all functions in this class of MaxEnt solutions is determined by Eq. (68).* The aforementioned different time evolutions of conditional probability distribution do not represent actual microscopic time evolutions. All that conditional probability distributions in this class of MaxEnt solutions represent is a statistical inference that predicts possible time evolutions from incomplete information about microscopic dynamics, represented by the set $\Omega(M)$ of phase space paths and the macroscopic constraint (63). By definition, the set M of possible microstates in Γ is invariant to Hamiltonian motion. Due to Eq. (58), MaxEnt conditional probability distributions satisfying the macroscopic constraint (63) predict only time evolutions entirely within the invariant set M. This description is compatible with the values of the constants of motion that shape the invariant set M. It takes into account the conservation laws that the Hamiltonian system has and predicts only the behavior and final states consistent with these conservation laws. In this way, description also incorporates memory effects as characteristic features of irreversible processes.

From the normalization property (40) of the conditional probability distribution given at time t_a by Eq. (68) one obtains the relation:

$$W(M) \exp\left[-\lambda_2((q_0, p_0)_\omega, t_0; t_a)\right] = 1, \tag{69}$$

with $W(M)$ as the measure, i.e., phase space volume of the invariant set M. Eq. (69) implies the following relation for Lagrange multiplier $\lambda_2((q_0, p_0)_\omega, t_0; t_a)$ at time t_a:

$$\lambda_2((q_0, p_0)_\omega, t_0; t_a) = \lambda_2(t_a). \tag{70}$$

Microstate probability distribution $f(q, p, t_a)$ at time t_a is then calculated using Eqs. (23) and (29), from the MaxEnt conditional probability distribution $D(q, p, t_a|(q_0, p_0)_\omega, t_0)$ at time t_a and the path probability distribution $F((q_0, p_0)_\omega, t_0)$ at initial time t_0. With the help of Eqs. (68) and (70) and the normalization property of the path probability distribution $F((q_0, p_0)_\omega, t_0)$, one then obtains

$$f(q, p, t_a) = \exp\left[-\lambda_2(t_a)\right]. \tag{71}$$

From Eqs. (68-71) it follows that, at time t_a, the MaxEnt conditional probability distribution and the corresponding microstate probability distribution are equal,

$$D(q, p, t_a|(q_0, p_0)_\omega, t_0) = f(q, p, t_a) = \exp\left[-\lambda_2(t_a)\right] = \frac{1}{W(M)}. \tag{72}$$

From Eq. (72) follows that possible microstates at time t_a in the invariant set M and possible phase space paths at time t_0 in the set $\Omega(M)$ are statistically uncorrelated, i.e., statistically independent. Eqs. (68-72) refer to phase space paths passing through infinitesimal surface elements around the set of points $\{(q_0, p_0) \mid (q_0, p_0) \in S_0(M), F((q_0, p_0)_\omega, t_0) \neq 0\}$ on the surface $S_0(M)$, as explained in Section III, Eqs. (28) and (29).

VI. SUMMARY AND CONCLUSIONS

From Eq. (72) and the definition of conditional information entropy $S_I^{DF}(t, t_0)$ in Eq. (34) and information entropy $S_I^f(t)$ in Eq. (37), one obtains their values at time t_a. The values of $S_I^{DF}(t_a, t_0)$ and $S_I^f(t_a)$ at time t_a are equal

$$S_I^f(t_a) = S_I^{DF}(t_a) = \log W(M). \tag{73}$$

In accordance with Eq. (39), it follows that Eq. (73) is possible if (and only if) possible phase space paths at time t_0 in the set $\Omega(M)$ and possible microstates at time t_a in the invariant set M are statistically independent. Interpretation of Eq. (73), in accordance with Eq. (39) and properties described by it, is a central result of the second approach to MaxEnt inference algorithm and leads to the following conclusion:

Maximization of the conditional information entropy $S_I^{DF}(t_a, t_0)$ at time t_a, subject to Liouville equation (41) regarded as a macroscopic constraint on time evolution, results in a maximum possible value of $S_I^{DF}(t_a, t_0)$. In reproducible time evolution of a closed system, information entropy $S_I^f(t)$ of the microstate probability distribution has its maximum value consistent with constraints. The values of $S_I^{DF}(t_a, t_0)$ and $S_I^f(t_a)$ at time t_a are equal (Eq. 73). This last property is possible if (and only if) the initial phase space paths and final microstates are statistically independent, i.e., uncorrelated. If (and only if) statistical independence is realized, conditional information entropy $S_I^{DF}(t_a, t_0)$ attains its upper bound given by $S_I^f(t_a)$. Therefore, if MaxEnt inference algorithm is applied on macroscopic time evolution of a closed Hamiltonian system as done here in second approach, maximization of the conditional information entropy $S_I^{DF}(t_a, t_0)$ includes all above properties and is equivalent to loss of statistical correlation.

It is important to emphasize that in the second approach, $S_I^{DF}(t_a, t_0)$ is maximized, subject to constraints (47) and (63), for some particular time t_a, such that $t_a > t_0$. In general, property of macroscopic systems is that they appear to randomize themselves between observations, provided that the observations follow each other by a time interval longer then a certain characteristic time τ called the relaxation time [17]. In the interpretation given here, relaxation time τ for a closed Hamiltonian system represents a characteristic time required for the described loss of correlation between the initial phase space paths and final microstates. It can also be said that τ represents a time interval during which predictions, based on incomplete information about possible microscopic time evolutions, become uncertain to a maximum extent compatible with the macroscopic data. This interpretation follows from Shannon's concept of information entropy as a measure of information and uncertainty [4], and from Jaynes formulation of predictive statistical mechanics [8]. It is underlined with the fact that in first approach, equivalent to complete knowledge of microscopic dynamics, $S_I^{DF}(t, t_0)$ is constant for all times $t \geq t_0$.

This interpretation is reflected in the role of Lagrange multipliers $\lambda_1((q_0, p_0)_\omega, t_0; t)$ and $\lambda_2((q_0, p_0)_\omega, t_0; t)$. They are required to satisfy Eq. (66), and by integrating it, one obtains the following relation:

$$\lambda_2((q_0, p_0)_\omega, t_0; t) = \int_{t_0}^{t} \lambda_1((q_0, p_0)_\omega, t_0; t')dt' + \lambda_2((q_0, p_0)_\omega, t_0; t_0), \tag{74}$$

for all t in the interval $t_0 \leq t \leq t_a$ of time evolution. Using Eqs. (69), (70), (73) and (74) one then obtains

$$S_I^f(t_a) = S_I^{DF}(t_a, t_0) = \log W(M) = \int_{t_0}^{t_a} \lambda_1((q_0, p_0)_\omega, t_0; t)dt + \lambda_2((q_0, p_0)_\omega, t_0; t_0). \tag{75}$$

It is clear from Eqs. (70), (74) and (75) that the value of Lagrange multiplier $\lambda_2((q_0, p_0)_\omega, t_0; t_a) \equiv \lambda_2(t_a)$ at time t_a is determined by the measure $W(M)$ of the invariant set M of possible microstates, i.e., by the volume of accessible phase space. Subsequent applications of the MaxEnt algorithm of described type for a closed Hamiltonian system without introducing additional constraints, should all result in increase of the volume of accessible phase space. From Eqs. (70), (74) and (75) it is then deduced that

$$\lambda_2(t_a) \geq \lambda_2(t_0). \tag{76}$$

Eq. (76) follows because in the described subsequent applications of MaxEnt, in strictly mathematical sense there are no other upper bounds on volumes of accessible phase space, except those given by the values of conserved quantities

and the conservation laws that the closed Hamiltonian system has. This restriction of the set of possible microstates gives an upper bound on volumes of accessible phase space and represents information that contains what is known about the structure of possible microstates. If volumes of accessible phase space are subsequently increased, the information about possible microstates is lost gradually along with our ability to predict the time evolution from incomplete information. The idea that irreversibility is related to gradual loss of information has been developed by Jaynes in the density-matrix formalism [2]. According to Jaynes [6], the real reason behind the second law, since phase space volume is conserved in the dynamical evolution (by Liouville's theorem), is a fundamental requirement on any reproducible process that the phase space volume W' compatible with the final state can not be less than the phase space volume W_0 which describes our ability to reproduce the initial state. In the limit of large number particles, Boltzmann's original interpretation of entropy of a macrostate as the logarithm of the classical phase space volume consistent with it, i.e., number of microstates by which the macroscopic state can be realized, is compatible with the Gibbs' formalism and the maximum-entropy inference. By considering W as the volume of the high-probability region in phase space, entropy maximization for all practical purposes finds the probability distribution which defines the largest high-probability region while agreeing with data [6, 7]. Macroscopic behavior that is predicted can be realized in the greatest number of ways compatible with data.

The values of $S_I^{DF}(t_a, t_0)$ and $S_I^f(t_a)$ at time t_a in Eq. (75) are equal to the maximum value of the *Boltzmann-Gibbs entropy*, compatible with the information about the structure of possible microstates. The Lagrange multiplier $\lambda_1((q_0, p_0)_\omega, t_0; t)$ integrated over time $t_0 \leq t \leq t_a$ in Eq. (75) is determined by the rate at which the macroscopic state characterized by the maximum value of Boltzmann-Gibbs entropy is reached in a reproducible time evolution. Integral over time in Eq. (75) and the quantity $\lambda_1((q_0, p_0)_\omega, t_0; t)$ can be identified with the change in entropy and the *entropy production* for a closed Hamiltonian system, respectively. If the information about microscopic dynamics of closed Hamiltonian system is considered complete, whether entropy production can be defined without recourse to coarse graining procedures, or macroscopic, phenomenological approaches, remains an open question. This opinion is underlined with the result given by Eq. (62) which implies that in the *objective* approach that assumes complete knowledge of microscopic dynamics, entropy produced by all individual phase space paths is equal to zero. In the approach that takes into account the incomplete nature of information about microscopic dynamics, concepts of macroscopic reproducibility and information loss become fundamental to the interpretation and description of macroscopic time evolution and irreversibility. This approach leads to a simple definition for entropy production.

[1] Jaynes, E.T.: Information Theory and Statistical Mechanics. Phys. Rev. 106, 620-630 (1957)
[2] Jaynes, E.T.: Information Theory and Statistical Mechanics. II. Phys. Rev. 108, 171-190 (1957)
[3] Gibbs, J.W.: Elementary Principles in Statistical Mechanics. Yale University Press, New Haven (1902)
[4] Shannon, C.E.: A Mathematical Theory of Communication. Bell Syst. Tech. J. 27, 379-423, 623-656 (1948). Reprinted In: Shannon C.E., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press, Urbana (1949)
[5] Jaynes, E.T.: Information Theory and Statistical Mechanics. In: Ford, K.W. (ed.) 1962 Brandeis Lectures in Theoretical Physics, vol. 3, pp. 181-218. W. A. Benjamin, Inc., New York (1963)
[6] Jaynes, E.T.: Gibbs vs Boltzmann Entropies. Am. J. Phys. 33, 391-398 (1965)
[7] Jaynes, E.T.: Where do we stand on Maximum Entropy?. In: Levine, R.D., Tribus, M. (eds.) The Maximum Entropy Formalism, pp. 15-118. M. I. T. Press, Cambridge, MA (1979)
[8] Jaynes, E.T.: The Minimum Entropy Production Principle. Ann. Rev. Phys. Chem. 31, 579-601 (1980)
[9] Jaynes, E.T.: Macroscopic Prediction. In: Haken, H. (ed.) Complex Systems - Operational Approaches in Neurobiology, Physics, and Computers, pp. 254-269. Springer-Verlag, Berlin (1985)
[10] Grandy, W.T.: Principle of Maximum Entropy and Irreversible Process. Phys. Rep. 62, 175-266 (1980)
[11] Garrod, C.: Statistical Mechanics and Thermodynamics. Oxford University Press, New York (1995)
[12] Jaynes, E.T.: The Second Law As Physical Fact and As Human Inference. Unpublished manuscript (1990). http://bayes.wustl.edu/etj/node2.html
[13] Zurek, W.H.: Algorithmic Randomness and Physical Entropy. Phys. Rev. A 40, 4731-4751 (1989)
[14] Khinchin, A.I.: Mathematical Foundations of Statistical Mechanics. Dover Publications, Inc., New York (1949)
[15] Goldstein, H.: Classical Mechanics, 2nd edn. Addison-Wesley, Reading (1980)
[16] Wan, F.Y.M.: Introduction to the Calculus of Variations and Its Applications. Chapman & Hall, New York (1995)
[17] Kittel, C.: Elementary Statistical Physics. John Wiley & Sons, Inc., New York (1958)