H-Theorems from Autonomous Equations

Wojciech De Roeck[1] , **Christian Maes**[2]
Instituut voor Theoretische Fysica, K.U.Leuven (Belgium)

Karel Netočný[3]
Institute of Physics AS CR, Prague

Abstract: The H-theorem is an extension of the Second Law to a time-sequence of states that need not be equilibrium ones. In this paper we review and we rigorously establish the connection with macroscopic autonomy.
If for a Hamiltonian dynamics for many particles, at all times the present macrostate determines the future macrostate, then its entropy is non-decreasing as a consequence of Liouville's theorem. That observation, made since long, is here rigorously analyzed with special care to reconcile the application of Liouville's theorem (for a finite number of particles) with the condition of autonomous macroscopic evolution (sharp only in the limit of infinite scale separation); and to evaluate the presumed necessity of a Markov property for the macroscopic evolution.

KEY WORDS: H-theorem, entropy, irreversible equations

1 Introduction

The point of the present paper is to make mathematically precise the application of Liouville's theorem in microscopic versions or derivations of the Second Law, under the assumption that an autonomous evolution is verified for the macroscopic variables in question. Microscopic versions of the Second Law, or perhaps more correctly, generalizations of the Second Law to nonequilibrium situations, are here referred to as H-Theorems.

The stability of points of a dynamical system can be demonstrated with the help of Lyapunov functions. Yet in general these functions are hard to find — there does not exist a construction or a general algorithm to obtain them. On the other hand, when the differential equation has a natural interpretation, as with a specific physical origin, we can hope to improve on trial and error. Think of the equations of irreversible thermodynamics where some approach to equilibrium is visible or at least expected. Take for example the diffusion equation

$$\frac{\partial n_t(r)}{\partial t} + \nabla \cdot J_r(n_t) = 0 \tag{1.1}$$

for the particle density $n_t(r)$ at time t and at location r in some closed box. That conservation equation is determined by the current J_r depending on the

[1]Aspirant FWO, U.Antwerpen

[2]email: `christian.maes@fys.kuleuven.be`

[3]K.N. acknowledges the support from the project AVOZ10100520 in the Academy of Sciences of the Czech Republic.

particle density via the usual phenomenology

$$\begin{aligned} J_r(n_t) &= \frac{1}{2}\chi(n_t(r))\,\nabla s'(n_t(r)) \\ &= -\frac{1}{2}D(n_t(r))\,\nabla n_t(r)) \end{aligned}$$

Here $D(n_t(r))$ is the diffusion matrix, connected with he mobility matrix χ via

$$\chi(n_t(r))^{-1}D(n_t(r)) = -s^{''}(n_t(r))\ \mathrm{Id}$$

for the identity matrix Id and the local thermodynamic entropy s. From irreversible thermodynamics the entropy should be monotone and indeed it is easy to check that $\int dr\, s(n_t(r))$ is non-decreasing in time t along (1.1).

Such equations and the identifications of monotone quantities are of course very important in relaxation problems. A generic relaxation equation is that of Ginzburg-Landau. There an order parameter m is carried to its equilibrium value via

$$\frac{dm}{dt} = -D\frac{\delta\Phi}{\delta m}$$

where D is some positive-definite operator, implying

$$\frac{d\Phi}{dt} \leq 0$$

for $\Phi(m)$ for example the Helmholtz free energy.
That scenario can be generalized. We are given a first order equation of the form

$$\frac{dm_t}{dt} = F(m_t), \quad m_t \in \mathbb{R}^\nu \tag{1.2}$$

with solution $m_t = \phi_t(m)$. It is helpful to imagine extra "microscopic" structure. One supposes that (1.2) results from a law of large numbers in which m_t is the macroscopic value at time t and ϕ_t gives its autonomous evolution. At the same time, there is an entropy $H(m_t)$ associated to the macroscopic variable and one hopes to prove that $H(m_t) \geq H(m_s)$ for $t \geq s$. That will be explained and mathematically detailed starting with Section 3.

Usually however, from the point of view of statistical mechanics, the problem is posed in the opposite sense. Here one looks for microscopic versions and derivations of the Second Law of thermodynamics. One starts from a microscopic dynamics and one attempts to identify a real quantity that increases along a large fraction of trajectories. We will show that such an H-theorem is valid for the Boltzmann entropy when it is defined in terms of these macroscopic observables that satisfy an autonomous equation (Propositions 3.1 en 3.2).

The heuristics is simple: when there is an autonomous deterministic evolution taking macrostate M_s at time s to macrostate M_t at time $t \geq s$, then, under the Liouville dynamics U the volume in phase space $|M_s| = |UM_s|$ is preserved. On the other hand, since about every microstate x of M_s evolves under U into one corresponding to M_t, we must have, with negligible error that $UM_s \subset M_t$. We conclude that $|M_s| \leq |M_t|$ which gives monotonicity of the

Boltzmann entropy $S = k_B \log |M|$.
That key-remark has been made before, most clearly for the first time on page 84 in [5], but see also e.g. [3] page 9–10, [4] page 280–281, [6] Fig.1 page 47, [10] page 278, page 301, and most recently in [8, 7]. We believe however it helps to add some mathematical precision here. For example, Liouville's theorem employs a finite number of particles and the autonomy of the macroscopic equation is probably only satisfied in some hydrodynamic limit where also the number of particles goes to infinity.

The set-up we start from here is a classical dynamical system and we show in what sense one can say that when a collection of variables obtains an autonomous evolution, the corresponding entropy will be monotone.

The following section gives context and motivation. We specify more clearly what we mean by an H-theorem and how it relates to the Second Law. The mathematics starts from Section 3.

2 Second Law derivations

For a thermodynamically reversible process in an equilibrium system with energy U and particle number N in a volume V, the entropy $S(U,N,V)$ as defined by Clausius satisfies

$$dS = \frac{1}{T}(dU + p\,dV - \mu\,dN)$$

where p is pressure, μ is chemical potential and T denotes the temperature. Following up on the reflections by Sadi Carnot about maximal efficiencies for heat engines, Clausius put forward the general law that the entropy of the universe is increasing. That so called Second Law is part of every course in thermodynamics but has a reputation of being hard to make mathematical sense of. With the words of Arnold *every mathematician knows it is impossible to understand an elementary course in thermodynamics*, [1] page 163. Also historians of science like Truesdell, [11] page 335, speaking about Clausius' arguments, declare *I cannot explain what I cannot understand.* The mystery surrounding the Second Law has been tried to be solved by many over the previous century. It would take us too far to review these attempts. We pick an argument that goes back to Gibbs.

2.1 Variational principle

Suppose that at time $t = 0$ a system of N particles is in complete thermal equilibrium at inverse temperature β. The Gibbs canonical distribution $\rho_\beta \sim \exp -\beta H_0$ represents all its macroscopic properties. Modulo some constants the entropy then equals

$$S(0) = -\int \rho_\beta \, \log \rho_\beta \, dx \tag{2.1}$$

with $dx \equiv dq_1 \dots dq_N dp_1 \dots dp_N$ the Liouville element over all particle coordinates and momenta. Suppose the system now undergoes an evolution in which no heat is flowing to or from the environment. The N-particle distribution $\rho(t)$

at time t varies according to the Liouville equation

$$\dot{\rho}(t) = \{H(t), \rho(t)\}$$

with $\rho(0) = \rho_\beta$ and time-dependent Hamiltonian $H(t), H(0) = H_0$. As a consequence,

$$S(0) = -\int \rho(t)\, \log \rho(t)\, dx$$

remains constant in time. At time $t = \tau$, we have as average energy

$$U_\tau = \int H(\tau)\, \rho(\tau)\, dx$$

and we can consider the Gibbs distribution ρ' for $H_\tau \equiv H(t = \tau)$ that maximizes the entropy functional

$$h(\rho) \equiv -\int \rho\, \log \rho\, dx, \quad U_\tau = \int H_\tau\, \rho\, dx \tag{2.2}$$

subject to the constraint that it has the average energy U_τ. ρ' is an equilibrium distribution with entropy

$$S(\tau) = h(\rho')$$

and, clearly, for all τ no matter how large or small,

$$S(\tau) = h(\rho') \geq h(\rho(\tau)) = h(\rho(0)) = S(0)$$

If therefore, say after fixing $H(t) = H_\tau, t \geq \tau > 0$, a new equilibrium gets installed in the sense that macroscopic quantities are no longer varying, the final equilibrium entropy will not be smaller than that of the initial equilibrium. That in a nutshell is the argument written down by Jaynes in 1965, Section IV of [5], for the validity of the Second Law of thermodynamics. It treats a specific case but the idea is very general. Before we criticize the argument in Sections 2.1.1–2.1.3, let us review one more scenario and since it makes no real difference, we take the opportunity to consider a quantum set-up.

The system is described by a finite-dimensional Hilbert space $\mathcal{H}$ for which one chooses a suitable orthogonal decomposition

$$\mathcal{H} = \bigoplus \mathcal{H}_\alpha$$

into linear subspaces. Denote by d_α the dimension of $\mathcal{H}_\alpha$ and let P_α be the projection on $\mathcal{H}_\alpha = P_\alpha \mathcal{H}$. We think of macrostates as labeled by the running index α. Initially the system is in constrained microcanonical equilibrium as described by the density matrix

$$\rho_{\hat{\mu}} \equiv \sum_\alpha \frac{\hat{\mu}(\alpha)}{d_\alpha} P_\alpha$$

for a given macro-statistics $\hat{\mu}(\alpha) \geq 0, \sum_\alpha \hat{\mu}(\alpha) = 1$. Its entropy is

$$S(0) = -\,\mathrm{Tr}[\rho_{\hat{\mu}}\, \log \rho_{\hat{\mu}}] \tag{2.3}$$

The quantum dynamics with Hamiltonian H is switched on and the density matrix evolves into $\rho(t)$ following the von Neumann equation

$$\dot{\rho}(t) = -i[H, \rho(t)]$$

be it backward or forward in time. That unitary dynamics preserves

$$S(0) = h(t) \equiv -\operatorname{Tr}[\rho(t)\,\log \rho(t)] \tag{2.4}$$

At time τ we look at the new macro statistics

$$\hat{\mu}_\tau(\alpha) = \operatorname{Tr}[P_\alpha\, \rho_\tau]$$

and we obtain the new equilibrium entropy as

$$S(\tau) = \sup_{p(\rho)=\hat{\mu}_\tau} -\operatorname{Tr}[\rho\,\log\rho]$$

where we take the supremum over all density matrices ρ with

$$p(\rho)(\alpha) \equiv \operatorname{Tr}[P_\alpha\, \rho] = \hat{\mu}_\tau(\alpha)$$

As before and for the same reason, $S(\tau) \geq S(0)$.

The previous arguments reduce the Second Law to inequalities between the entropies of equilibrium states. That Second Law of Thermodynamics finds Gibbs variational principle as counterpart in statistical mechanics. It can be used to compare the entropy of the initial and the final state of the universe when some system has passed from one equilibrium state to a second one. Equilibrium is described via a variational principle, maximizing the entropy subject to constraints. When a constraint is lifted, even in a time-dependent way, new macrostates can be explored and the entropy will increase until a new equilibrium is installed. There are however various weaknesses in such an approach as we now specify.

2.1.1 Stone-wall character

Note that the mathematical arguments above remain valid and the conclusion $S(\tau) \geq S(0)$ remains true even in the case when the universe is small or contains only few particles. Indeed, also within the thermodynamic formalism in the theory of dynamical systems one meets the very same variational principle, equilibrium states and notions of entropy and at no moment is there a restriction that the system should be large. Yet for small systems, there is no Second Law, at least not in its stone-wall character as manifested in thermodynamics. For small systems, fluctuations can be relatively large. Thermodynamics speaks about average values but with the understanding that these averages correspond to *typical* values, and that only works when considering macroscopic quantities. As such have the functionals (2.2) or (2.4) no relation with the thermodynamic entropy. Only for macroscopic equilibrium systems do (2.1) or (2.2) acquire the Boltzmann-interpretation as measure of the volume of phase space corresponding to the macroscopic constraint, $S = k_B \log W$. The law of large numbers then ensures that the volume in phase space that corresponds to the equilibrium values of the macroscopic quantities is so very much larger than the volume of nonequilibrium parts. A thermodynamic entropy difference $s' - s$ per particle of the order of Boltzmann's constant k_B is physically reasonable and corresponds to a total reversible heat exchange $T(S' - S) = NT(s' - s)$ of about 0.1 millicalorie at room temperature T and to a phase volume ratio of

$$\frac{W}{W'} = \exp -(S' - S)/k_B = e^{-10^{20}}$$

when the number of particles $N = 10^{20}$.

2.1.2 Time-symmetry

The arguments of Section 2.1 work equally well for $\tau < 0$ as for $\tau > 0$, backward or forward in time. In that way, these versions of the Second Law cannot be related to the infamous Arrow of Time. Even worse, the door is open for all kinds of Loschmidt constructions. Clearly, the argument goes, one can consider the system in its final macroscopic state and invert all the momenta; the dynamically reversible time-evolution takes the system back to what was originally the initial state having lower entropy.
What is missing here are considerations of the initial state and a notion of what typical states could be. Again, the conclusion above that $S(\tau) \geq S(0)$ is true but seems to give no insight in the relation with thermodynamic time and macroscopic irreversibility.

2.1.3 Initial states

As formulated above, the Second Law considers very special initial and final states. They are equilibrium states. Note in particular, that the argument says nothing about monotonicity of entropy — we cannot infer e.g. that $S(\tau) \geq S(\tau/2)$. The question therefore arises whether the Second Law has also a validity outside equilibrium. What if initially a bomb explodes and brings the system way out of equilibrium. In the course of relaxation to equilibrium, can we then still speak about an increasing entropy. It certainly looks irreversible.

From the above, we start to recognize what is missing. We need a microscopic version of the Second Law which would allow us to bring in details of the initial state and for which the evolution is monitored on the level of the micro states. At the same time, we can hope for what is not uncharacteristic to various mathematical projects: by proving something stronger we take care of all seeming paradoxes that overshadowed a true but much weaker statement. That is exactly what Boltzmann achieved in 1872 in the derivation and study of the Boltzmann equation. That stronger statement is called an H-theorem and it is very much linked with macroscopic reproducibility or, more mathematically, with the existence of an autonomous equation for the macroscopic variables.
A somewhat formal and abstract argument can be presented as follows.

2.2 H-theorem

The word H-theorem originates from Boltzmann's work on the Boltzmann equation and the identification of the so called H-functional. The latter plays the role of entropy for a dilute gas and is monotone along the Boltzmann equation.

One often does not distinguish between the Second Law and the H-theorem. Here we do (and the entropy will from now on be denoted by the symbol H, not be confused with the Hamiltonian of Section 2.1).

We start by repeating the heuristics at the end of the Introduction in a somewhat more abstract fashion. Consider a transformation f on states x of a measure space (Ω, ρ). The measure ρ is left invariant by f. Suppose there is a sequence $(M_n), n = 0, 1, \dots$ of subsets $M_n \subset \Omega$ for which

$$\rho((f^{-1}M_{n+1})^c \cap M_n) = 0, \quad n = 0, \dots, \tau \tag{2.5}$$

In other words, M_{n+1} should contain about all of the image fM_n. Then,

$$\rho(M_{n+1}) = \rho(f^{-1}M_{n+1}) \geq \rho(f^{-1}M_{n+1} \cap M_n) = \rho(M_n)$$

and $\rho(M_n)$ or $\log \rho(M_n)$ is non-decreasing for $n = 0, \dots, \tau$.
We now add more structure to realize that starting from a more microscopic level.

Imagine there is a map α from Ω to a class of subsets of Ω for which

$$M_n(x) = \alpha(f^n x)$$

satisfies (2.5) for all $x \in \Omega'$, a subset of Ω. We write $H(y) \equiv \log \rho(\alpha(y))$. Then, changing notation to $x_n = f^n x$,

$$H(x_n) = \log \rho(\alpha(x_n)) \tag{2.6}$$

is monotone in $n = 0, \dots, \tau$. In other words the entropy (2.6) is monotone along the paths starting from states in Ω'.
The word entropy presupposes some interpretation here. One can imagine that the (M_n) are selected from phase space regions that correspond to some macroscopic state. Each M_n is a particular macroscopic state.
The condition (2.5) basically requires that from the knowledge of $\alpha(x), x \in \Omega'$ we can predict all $\alpha(x_n), n = 0, \dots, \tau$. Therefore the transformation f gets replaced on the $\alpha-$level with a new autonomous dynamics. One must add here a warning however. One can imagine a macroscopic dynamics satisfying only

$$\rho((f^{-n}M_n)^c \cap M_0) = 0 \qquad n = 0, \dots, \tau \tag{2.7}$$

i.e., the initial macrostate M_0 determines the whole trajectory as well but the macrodynamics is possibly not Markovian as it was in (2.5). Loosely speaking, that can happen when almost all of a macrostate M_1 is mapped into macrostate M_2, and nearly all of M_2 is mapped into M_3, but $fM_1 \subset M_2$ is not typically mapped into M_3.
We have in mind an order parameter m_t whose evolution is like that of a damped oscillator and which fluctuates with decreasing amplitude around its asymptotic value: $m_t = m_0 r^t \cos \omega t, |r| < 1$. An example can be found in Section 3.3 of [2]. Then, under (2.7), one only obtains that $\rho(M_n) \geq \rho(M_0)$ but the $\rho(M_n)$ are allowed to oscillate as a function of n.
Hence, it is crucial that the macroscopic dynamics is autonomous in the Markovian sense (2.5), in order to get the usual H-theorem. To get an H-theorem for

the more general situation (e.g. for higher order differential equations), the entropy (2.6) would have to be generalized. In other words, condition (2.5) of autonomy goes hand in hand with the interpretation of (2.6) as an entropy.

For a system composed of many particles, we can expect a (Markovian) autonomous evolution over a certain time-scale for certain macroscopic variables. These are indexed by the map α. In that case (2.6) coincides with the Boltzmann entropy: it calculates the volume in phase space compatible with some macroscopic constraint (like fixing energy and some density- or velocity profile). The identification with thermodynamic entropy (in equilibrium) and with expressions like (2.1) or (2.3) then arises from considerations of equivalence of ensemble. In a way, the H-theorem is a nonequilibrium version of the Second Law — not only considering initial and final equilibria but also the entropy of the system as it evolves possibly away from equilibrium.

All of what follows concentrates on mathematically precise and physically reasonable formulations of (2.5) and (2.6) to obtain monotonicity of entropy. One should indeed not forget that autonomy is mostly expected in some scaling limits, e.g. as the number of particles N goes to infinity etc. In that case, the dynamical system must be parameterized with N and (2.5) is only valid in some limit $N \uparrow \infty$. The main purpose is therefore to clarify a theoretical/mathematical question; not to include new results for specific models. The only difficulty is to identify the appropriate set of assumptions and definitions; from these the mathematical arguments will be relatively short and easy but they clarify the whole conceptual set-up.

3 Classical dynamical systems

Let N be an integer, to be thought of as the number of degrees of freedom or as a scaling parameter, that indexes the dynamical system $(\Omega^N, U_t^N, \rho^N)$. Ω^N is the phase space with states $x \in \Omega^N$ and is equipped with a probability measure ρ^N, invariant under the dynamics $U_t^N : \Omega^N \to \Omega^N$.

We suppose a map

$$m^N : \Omega^N \to \mathcal{F} \tag{3.1}$$

which maps every state x into an element $m^N(x)$ of a metric space $(\mathcal{F}, d)$ (independent of N). For $\mathcal{F}$ one can have in mind $\mathbb{R}^n$ for some integer n or a space of real-valued functions on a subset of $\mathbb{R}^n$, with the interpretation that $m^N(x)$ gives the macroscopic state corresponding to the microscopic state x. For $m, m' \in \mathcal{F}$ and $\delta > 0$ we introduce the notation $m' \overset{\delta}{=} m$ for $d(m', m) \leq \delta$.

3.1 Infinite scale separation

We start here by considering the limit $N \uparrow +\infty$. In that limit the law of large numbers starts to play with deviations governed by

$$H(m) \equiv \lim_{\delta \downarrow 0} \lim_{N \uparrow +\infty} \frac{1}{N} \log \rho^N\big(m^N(x) \overset{\delta}{=} m\big), \quad m \in \mathcal{F} \tag{3.2}$$

That need not exist in general, but we make that definition part of our assumptions and set-up. For what follows under Proposition 3.1 it is in fact sufficient to take the lim sup in (3.2) (if we also take the lim sup in the next (3.3)) but for simplicity we prefer here to stick to the full limit. The limit (3.2) is then a natural notion of entropy à la Boltzmann.

The macroscopic observables are well-chosen when they satisfy an autonomous dynamics, sharply so in the proper limit of scales. Here we assume dynamical autonomy in the following rather weak sense: there is an interval $[0,T]$ and a map $\phi_t : \mathcal{F} \to \mathcal{F}$ for all $t \in [0,T]$ such that $\forall m \in \mathcal{F}, \forall \delta > 0$, and $0 \leq s \leq t \leq T$

$$\lim_{\kappa \downarrow 0} \lim_{N \uparrow +\infty} \frac{1}{N} \log \rho_N \Big(m^N(U_t^N x) \stackrel{\delta}{=} \phi_t(m) \big| m^N(U_s^N x) \stackrel{\kappa}{=} \phi_s(m) \Big) = 0 \qquad (3.3)$$

Proposition 3.1. *$\forall m \in \mathcal{F}$ and for all $0 \leq s \leq t \leq T$,*

$$H(\phi_t(m)) \geq H(\phi_s(m)) \qquad (3.4)$$

Proof. Writing out $H(\phi_t(m))$ we find that for every $\kappa > 0$

$$\begin{aligned} \log \rho^N \big(m^N(x) \stackrel{\delta}{=} \phi_t(m) \big) &= \log \rho^N \big(m^N(U_t^N x) \stackrel{\delta}{=} \phi_t(m) \big) \\ &\geq \log \rho^N \Big(m^N(U_t^N x) \stackrel{\delta}{=} \phi_t(m) \big| m^N(U_s^N x) \stackrel{\kappa}{=} \phi_s(m) \Big) \\ &+ \log \rho^N \big(m^N(U_s^N x) \stackrel{\kappa}{=} \phi_s(m) \big) \end{aligned} \qquad (3.5)$$

The equality uses the invariance of ρ^N and we can use that again for the last term in (3.5). We divide (3.5) by N and we first take the limit $N \uparrow +\infty$ after which we send $\kappa \downarrow 0$ and then $\delta \downarrow 0$. □

3.1.1 Semigroup property

If the dynamics (U_t^N) satisfies the semigroup property

$$U_{t+s}^N = U_t^N U_s^N \quad t, s \geq 0 \qquad (3.6)$$

and there is a unique macroscopic trajectory (ϕ_t) satisfying (3.3), then

$$\phi_{t+s} = \phi_t \circ \phi_s \quad t, s \geq 0 \qquad (3.7)$$

In practice the map ϕ_t will mostly be the solution of a set of first order differential equations.

3.1.2 Autonomy on a countable partition

It cannot really be seen from the heuristics but the limit procedure is necessary and physically relevant; the limiting procedure $N \uparrow +\infty$ in (3.3) is needed to reach the possibility that the entropy is *strictly* increasing. We take the same start as above (3.1) but we assume also that $U_t^N \Omega^N = \Omega^N$ and we specify (3.1) as follows: suppose there is a countable partition $(M_i), i \in I, I$ countable, $\rho^N(M_i) > 0$ of Ω for which there is autonomy in the sense that there is a map a_t on I with $\rho^N(U_t^N M_i \cap M_{a_t i}^c) = 0$. That means that $U_t^N M_i \subset M_{a_t i}$ up to

ρ^N-measure zero. Then, necessarily, a_t is a bijection on I. Suppose indeed for a moment that there is a $j \in I$ so that $j \neq a_t i$ for no matter which $i \in I$, or more precisely

$$\sum_{i \in I} \rho^N(U_t^N M_i \cap M_j) = 0$$

Then obviously $\rho^N(M_j) = 0$ which is excluded. Hence, a_t is a bijection on I. Together with the fact that ρ is normalized, that shows that the entropy is constant in time. All that shows the necessity of the limiting procedure in the partitioning of the phase space to allow for a relaxed notion of autonomy; the autonomy then only holds in the limit.
All that does not mean that the limit $N \uparrow +\infty$ is unavoidable. In fact, physically speaking, we are very much interested in a statement which is true for very large but finite N. Such a statement will come in the next section 3.2.

3.1.3 Reproducibility

The autonomy assumption (3.3) says that, in some limit, the manifest image of the system does no longer depend on the detailed properties of the previous microstate(s) but can be derived from knowing the past manifest condition. Jaynes, in [3, 4, 5], refers to it as macro-reproducibility, saying that the evolution on macroscale gets reproduced when initially the correct macroscopic condition was installed. One should perhaps not take that too literally; reproducibility refers to experiments and people doing the experiment while the property of having autonomous macro-behavior is of course independent of us all; it is true without observers.

3.1.4 Reversibility

Equation (3.3) invites the more general definition of a large deviation rate function for the transition probabilities

$$\begin{aligned} &-J_{t,s}(m, m') \equiv \qquad (3.8)\\ &\lim_{\delta \to 0} \lim_{\kappa \to 0} \lim_{N \to \infty} \frac{1}{N} \log \rho^N(m^N(U_t^N x) \stackrel{\delta}{=} m' \,|\, m^N(U_s^N x) \stackrel{\kappa}{=} m), \qquad t \geq s \end{aligned}$$

which we assume exists. The bounds of (3.5) give

$$H(m') \geq H(m) - J_{t,s}(m, m') \qquad (3.9)$$

for all $m, m' \in \mathcal{F}$ and $t \geq s$. In particular, quite generally,

$$H(\phi_t(m)) \leq H(\phi_s(m)) + J_{t,s}(\phi_t(m), \phi_s(m)), \qquad t \geq s \qquad (3.10)$$

while, as from (3.3), $J_{t,s}(\phi_s(m), \phi_t(m)) = 0$. On the other hand, if the dynamical system $(\Omega^N, U_t^N, \rho^N)$ is reversible under an involution π^N, $U_t^N = \pi^N U_{-t}^N \pi^N$ such that $\rho^N \pi^N = \rho^N$, $\pi^N m^N = m^N$, then

$$H(m') - J_{t,s}(m', m) = H(m) - J_{t,s}(m, m') \qquad (3.11)$$

for all $m, m', t \geq s$. Hence, under dynamical reversibility (3.10) is an equality:

$$J_{t,s}(\phi_t(m), \phi_s(m)) = H(\phi_t(m)) - H(\phi_s(m)), \qquad t \geq s \qquad (3.12)$$

Remarks on the H-theorem for irreversible dynamical systems have been written in [9].

3.1.5 Propagation of constrained equilibrium

The condition (3.3) of autonomy needs to be checked for all times $t \geq s \geq 0$, starting at time zero from an initial value m. Obviously, that condition is somehow related to – yet different from Boltzmann's Stosszahlansatz. The latter indeed corresponds more to the assumption that any initial constrained equilibrium state at time zero evolves to new constrained equilibria at times $t > 0$. Formally we think of the heuristics at the end of Section 2 and we consider a region M_0 in phase space corresponding to some macroscopic state and its image UM_0 after some time t. We then have in mind to ask that for all phase space volumes A

$$\frac{|UM_0 \cap A|}{|UM_0|} = \frac{|M_t \cap A|}{|M_t|}$$

which means that the evolution takes the equilibrium constrained with $x \in M_0$ to a new equilibrium at time t constrained at $x \in M_t$. To be more precise and turning back to the present mathematical context, we consider the following condition:

$$\lim_{\epsilon \downarrow 0} \lim_{N \uparrow \infty} \frac{1}{N} \log \frac{\rho^N\{x \in A_i^N \,|\, m^N(x) \stackrel{\epsilon}{=} \phi_t(m)\}}{\rho^N\{U_t^N x \in A_i^N \,|\, m^N(x) \stackrel{\epsilon}{=} m\}} = 0 \qquad (3.13)$$

for all $m \in \mathcal{F}$, $t > 0, i = 1, 2$, $A_1^N \equiv \{m^N(x) \stackrel{\delta}{=} \phi_t(m)\}$ and
$A_2^N \equiv \{m^N(U_t^N (U_s^N)^{-1} x) \stackrel{\delta}{=} \phi_t(m)\}$.
Arguably, (3.13) is a (weak) version of propagation of constrained equilibrium. We check that it actually implies condition (3.3), and hence the H-theorem. First, choosing $A_1^N = \{m^N(x) \stackrel{\delta}{=} \phi_t(m)\}$, (3.13) yields

$$\lim_{\epsilon \downarrow 0} \lim_{N \uparrow \infty} \frac{1}{N} \log \rho^N\{m^N(U_t^N x) \stackrel{\delta}{=} \phi_t(m) \,|\, m^N(x) \stackrel{\epsilon}{=} m\} = 0$$

Second, using the invariance of ρ^N,

$$\rho^N\{m^N(U_t^N x) \stackrel{\delta}{=} \phi_t(m) \,|\, m^N(U_s^N x) \stackrel{\epsilon}{=} \phi_s(m)\}$$
$$= \rho^N\{m^N(U_t^N x) \stackrel{\delta}{=} \phi_t(m) \,|\, m^N(x) \stackrel{\epsilon}{=} m\} \, \frac{\rho^N\{m^N(U_t^N (U_s^N)^{-1} x) \stackrel{\delta}{=} \phi_t(m) \,|\, m^N(x) \stackrel{\epsilon}{=} \phi_s(m)\}}{\rho^N\{m^N(U_t^N (U_s^N)^{-1} U_s^N x) \stackrel{\delta}{=} \phi_t(m) \,|\, m^N(x) \stackrel{\epsilon}{=} m\}}$$

and by applying condition (3.13) once more but now with
$A_2^N = \{m^N(U_t^N (U_s^N)^{-1} x) \stackrel{\delta}{=} \phi_t(m)\}$ and taking the limits, we get (3.3). (Note that we have actually also used here that U_t^N is invertible or at least that $U_t^N \circ (U_s)^{-1}$, $t \geq s$ is well defined.)

3.2 Pathwise

Consider

$$H^{N,\epsilon}(m) = \frac{1}{N} \log \rho^N\{m^N(x) \stackrel{\epsilon}{=} m\} \tag{3.14}$$

for (macrostate) $m \in \mathcal{F}$. For a microstate $x \in \Omega^N$, define

$$\overline{H}^{N,\epsilon}(x) = \sup_{m \stackrel{\epsilon}{=} m^N(x)} H^{N,\epsilon}(m)$$

$$\underline{H}^{N,\epsilon}(x) = \inf_{m \stackrel{\epsilon}{=} m^N(x)} H^{N,\epsilon}(m)$$

Instead of the hypothesis (3.3) of macroscopic autonomy, we assume here that

$$\lim_{\epsilon \downarrow 0} \lim_{N \uparrow \infty} \rho^N\{m^N(U_t^N x) \stackrel{\delta}{=} \phi_t(m) \,|\, m^N(U_s x) \stackrel{\epsilon}{=} \phi_s(m)\} = 1 \tag{3.15}$$

for all $m \in \mathcal{F}$, $\delta > 0$ and $0 < s < t$. That corresponds to the situation in (3.3) but where there is a unique macroscopic trajectory, which one observes typically. Now we have

Proposition 3.2. *Assume* (3.15). *Fix a finite sequence of times* $0 < t_1 < \ldots < t_K$. *For all* $m \in \mathcal{F}$, *there exists* $\delta_0 > 0$ *such that for all* $\delta \leq \delta_0$

$$\lim_{\epsilon \downarrow 0} \lim_{N \uparrow \infty} \rho^N\left[\overline{H}^{N,\delta}(U_{t_j}^N x) \geq \underline{H}^{N,\delta}(U_{t_{j-1}}^N x) - \frac{1}{N},\, j = 1, \ldots, K \,|\, m^N(x) \stackrel{\epsilon}{=} m\right] = 1 \tag{3.16}$$

Proof. Put

$$g^{N,\delta,\epsilon}(s,t,m) \equiv 1 - \rho^N\{m^N(U_t^N x) \stackrel{\delta}{=} \phi_t(m) \,|\, m^N(U_s x) \stackrel{\epsilon}{=} \phi_s(m)\} \tag{3.17}$$

then

$$\rho^N\left[m^N(U_{t_j} x) \stackrel{\delta}{=} \phi_{t_j}(m), j = 1, \ldots, K \,|\, m^N(x) \stackrel{\epsilon}{=} m\right] \geq 1 - \sum_{j=1}^K g^{N,\delta,\epsilon}(0, t_j, m) \tag{3.18}$$

Whenever $m^N(U_t x) \stackrel{\delta}{=} \phi_t(m)$, then

$$\underline{H}^{N,\delta}(U_t^N x) \leq H^{N,\delta}(\phi_t(m)) \leq \overline{H}^{N,\delta}(U_t^N x)$$

As a consequence, (3.18) gives

$$\rho^N\left[\underline{H}^{N,\delta}(U_{t_j} x) \leq H^{N,\delta}(\phi_{t_j}(m)) \leq \overline{H}^{N,\delta}(U_{t_j} x), j = 1, \ldots, K \,|\, m^N(x) \stackrel{\epsilon}{=} m\right]$$
$$\geq 1 - \sum_{j=1}^K g^{N,\delta,\epsilon}(0, t_j, m) \tag{3.19}$$

By the same bounds as in (3.5) we have

$$H^{N,\delta}(\phi_{t_j}(m)) \geq H^{N,\delta}(\phi_{t_{j-1}}(m)) + \frac{1}{N} \log[1 - g^{N,\delta,\delta}(t_{j-1}, t_j, m)]$$

The proof is now finished by choosing δ_0 such that for $\delta \leq \delta_0$ and for large enough N (depending on δ).

$$\min_{j=1}^{K} \big(\log[1 - g^{N,\delta,\delta}(t_{j-1}, t_j, m)] \big) \geq -1$$

□

3.3 Additional remarks

The above remains essentially unchanged for stochastic dynamics. Instead of the dynamical system (Ω, U_t, ρ) one considers any stationary process $(X_t^N)_{t \in \mathbb{R}_+}$ with the law $\mathbf{P}_N$; denote by ρ^N the stationary measure. The entropy is defined as in (3.2) but with respect to ρ^N. The (weak) autonomy in the sense of Section 3.1 is then

$$\lim_{\kappa \downarrow 0} \lim_{N \uparrow +\infty} \frac{1}{N} \log \mathbf{P}_N[m^N(X_t^N) \stackrel{\delta}{=} \phi_t(m) \big| m^N(X_s^N) \stackrel{\kappa}{=} \phi_s(m)] = 0$$

On the other hand some essential changes are necessary when dealing with quantum dynamics. The main reason is that, before the limit $N \uparrow +\infty$, macroscopic variables do not commute so that a counting or large deviation type definition of entropy is highly problematic. We keep the solution for a future publication.

While some hesitation or even just confusion of terminology and concepts have remained, the physical arguments surrounding an H-theorem have been around for more than 100 years. The main idea, that deterministic autonomous equations give an H-theorem when combined with the Liouville theorem, is correct but the addition of some mathematical specification helps to clarify some points. In this paper, we have repeated the following points

1. There is a difference between the Second Law of Thermodynamics when considering transformations between equilibrium states, and microscopic versions, also in nonequilibrium contexts, in which the Boltzmann entropy is evaluated and plays the role of an H-function.

2. The autonomy of the macroscopic equations should be understood as a Markov property (first order differences in time) and should not be confused with propagation of equilibrium. Mostly, that autonomy only appears sharply in the limit of infinite scale separation between the microscopic world and the macroscopic behavior. A specific limiting argument is therefore required to combine it with Liouville's theorem about conservation of phase space volume for finite systems.

As a final comment, there remains the question how useful such an analysis can be today. Mathematically, an H-theorem is useful in the sense of giving a Lyapunov function for a dynamical system, to which we alluded in the introduction. Physically, an H-theorem gives an extension and microscopic derivation of the Second Law of thermodynamics. One point which was however not mentioned here before, was much emphasized in years that followed Boltzmann's pioneering work, in particular by Albert Einstein. The point is that one can usefully turn the logic around. The statistical definition of entropy starts from a specific choice of microstates. If for that choice, the corresponding macroscopic evolution is not satisfying an H-theorem, then we know that our picture of the microstructure of the system may be inadequate. In other words, we can obtain information about the microscopic structure and dynamics from the autonomous macroscopic behavior. Then, instead of concentrating on the derivation of the macroscopic evolutions with associated H-theorem, we use the phenomenology to discover crucial features about the microscopic world. That was already the strategy of Einstein in 1905 when he formulated the photon-hypothesis.

References

[1] V. Arnold: *Contact geometry: the geometrical method of Gibbs' thermodynamics.* In: Proceedings of the Gibbs Symposium, D. Caldi and G. Mostow (eds.), Providence, American Mathematical Society, pp. 163–179 (1990).

[2] W. De Roeck, T. Jacobs, C. Maes, K. Netočný: *An Extension of the Kac ring model*, J. Phys. A: Math Gen. **36**, 11547–11559 (2003).

[3] E.T. Jaynes: *The second law as physical fact and as human inference* (1990) (unpublished), download from `http://www.etjaynescenter.org/bibliography.shtml`

[4] E.T. Jaynes: *The evolution of Carnot's principle.* In: Maximum-Entropy and Bayesian Methods in Science and Engineering, 1, G. J. Erickson and C. R. Smith (eds.), Kluwer, Dordrecht, pp. 267–281 (1988).

[5] E.T. Jaynes: *Gibbs vs Boltzmann entropies.* In: Papers on Probability, Statistics and Statistical Physics, Reidel, Dordrecht (1983). Originally in Am. J. Phys. **33**, 391 (1965).

[6] A.J.M. Garrett: *Irreversibility, probability and entropy*, From Statistical Physics to Statistical Inference and Back, P. Grassberger and J.-P. Nadal (eds.), pp. 47–75, Kluwer (1994).

[7] P.L. Garrido, S. Goldstein and J.L. Lebowitz: *The Boltzmann Entropy for Dense Fluids Not in Local Equilibrium.* Physical Review Letters **92**, 050602 (2003).

[8] S. Goldstein and J.L. Lebowitz: *On the (Boltzmann) Entropy of Nonequilibrium Systems.* Physica D (2004).

[9] C. Maes: *Fluctuation relations and positivity of the entropy production in irreversible dynamical systems*, Nonlinearity **17**, 1305–1316 (2004).

[10] C. Maes and K. Netočný: *Time-reversal and Entropy*, J. Stat. Phys. **110**, 269–310 (2003).

[11] C. Truesdell: The Tragicomical History of Thermodynamics 1822-1854, New York, Springer-Verlag (1980).