

# On the ergodic theorem and information loss in statistical mechanics

Andreas Henriksson

Received: date / Accepted: date

**Abstract** In this article, it is argued that, for a classical Hamiltonian system which is closed, the ergodic theorem emerge from the Gibbs-Liouville theorem in the limit that the system has evolved for an infinitely long period of time. In this limit, from the perspective of an ignorant observer, who do not have perfect knowledge about the complete set of degrees of freedom for the system, distinctions between the possible states of the system, i.e. the information content, is lost leading to the notion of statistical equilibrium where states are assigned equal probabilities. Finally, by linking the concept of entropy, which gives a measure for the amount of uncertainty, with the concept of information, the second law of thermodynamics is expressed in terms of the tendency of an observer to loose information over time.

**Keywords** Gibbs-Liouville theorem · Ergodic theorem · Statistical equilibrium · Second law of thermodynamics

**PACS** 05.20.-y · 05.70.-a · 02.50.-r

## Contents

1	The Gibbs-Liouville theorem . . . . .	1
2	Uncertainty and indistinguishability . . . . .	1
3	Conservation of classical probability . . . . .	2
4	Statistical equilibrium . . . . .	3
5	The ergodic theorem . . . . .	3
6	Microcanonical probability distribution . . . . .	3
7	Ergodicity breaking . . . . .	4
8	Is information conserved or lost? . . . . .	4
9	Entropy as a measure of uncertainty . . . . .	4
10	The second law of thermodynamics . . . . .	5

Andreas Henriksson  
St.Olav school, Jens Zetlitzgt. 33, 4008, Stavanger; Norway  
E-mail: andreas.henriksson@skole.rogfk.no  
ORCID: 0000-0001-9014-4320

## 1 The Gibbs-Liouville theorem

For classical Hamiltonian systems, the Gibbs-Liouville theorem [1][2] is equivalent to the statement that the divergence of the Hamiltonian phase flow velocity vanish, i.e. that

$$\nabla \cdot \mathbf{v} = 0 \quad (1)$$

where

$$\mathbf{v} = (\dot{q}, \dot{p}) = \left( \frac{\partial \mathcal{H}}{\partial p}, -\frac{\partial \mathcal{H}}{\partial q} \right) \quad (2)$$

where  $\mathcal{H}$  is the Hamiltonian of the system. In this article, the Gibbs-Liouville theorem is interpreted to represent a mathematical statement on the deterministic evolution of classical Hamiltonian systems, i.e. that distinctions between the possible states of the system, or, equivalently, the information content within the system, is conserved in time [3].

## 2 Uncertainty and indistinguishability

Even when ignoring the laws of quantum mechanics, which place a fundamental limit on the precision which can be gained, the dynamical evolution of a system is quite complicated. Most systems of interest contain a vast amount of particles that interact in complicated ways. For such large systems, it is usually very hard to track the individual evolution of each particle as the system evolve in time. Perfect knowledge about the position and velocity, or momenta, of each individual particle is lost. It is lost not because of a fundamental violation of the principle of information conservation but merely because of the difficulty for an observer to keep track of all the degrees of freedom. Therefore, from the perspective of the observer, there is an uncertainty  $\Delta q$

associated with the position of a state and an uncertainty  $\Delta p$  associated with the momentum of a state. For this reason, the observer is unable to determine with absolute certainty the state of the system at any given time. The observer can only determine whether or not the system occupy a state which lie within any given region  $\Omega_j$  on phase space, whose volume  $V_{\Omega_j}$  is given by the uncertainties  $\Delta q$  and  $\Delta p$ , i.e.

$$V_{\Omega_j} = \Delta q \Delta p \quad (3)$$

The volume  $V_{\Omega_j}$  is thus a measure of how ignorant the observer is about the details of the system, in the sense that the observer cannot locate an individual state to a greater precision than the size of  $\Omega_j$ . Due to this lack of precision, the observer is unable to distinguish between states that lie within  $\Omega_j$ . All states within  $\Omega_j$ , with their individual sets of degrees of freedom, has, from the perspective of the observer, collapsed into a single state whose single set of degrees of freedom is given by  $q + \Delta q$  and  $p + \Delta p$ . This so-called coarse-grained, or mixed, state is not a fundamental, or pure, state of the system. It is a description that average over all pure states within  $\Omega_j$ . Put differently, a mixed state  $\psi_j$ ,  $j \in [1, M]$ , where  $M$  is the number of mixed states on phase space, is a subjective representation, by an ignorant observer, of a collection of pure states  $\phi_\alpha$ ,  $\alpha \in [1, N]$ , where  $N$  is the number of pure states within  $\Omega_j$ . As the system evolve in time, the observer is only able to measure the coarse-grained flow, i.e. the jumping from one mixed state  $\psi_j$  to a different mixed state  $\psi_i$ ,  $i \neq j$ .

It should be noted that due to the lack of perfect knowledge about all the relevant degrees of freedom, the observer is unable to predict a unique evolutionary path on phase space along which the system evolve.

### 3 Conservation of classical probability

Due to the ignorance of the observer, i.e. the observers inability to distinguish the set of pure states within any given coarse-grained region  $\Omega_j$ , it is necessary to introduce the notion of probability on phase space. Let  $P_j$  be the probability that the system occupy  $\Omega_j$  and let  $P_\alpha$  be the probability that the system occupy the pure state  $\phi_\alpha$  within  $\Omega_j$ . If the observer know with absolute certainty that the system occupy the mixed state  $\psi_j$  and not some other state  $\psi_i$ ,  $i \neq j \in [1, M]$ , it is given that

$$P_i = 0, \forall i \neq j \in [1, M] \quad (4)$$

$$P_j \equiv \sum_{\alpha=1}^N P_\alpha = 1 \quad (5)$$

For continuous systems, the summation is replaced by an integral, i.e.

$$P_j \equiv \int_{\Omega_j} P_\alpha dV_\alpha = 1 \quad (6)$$

where  $dV_\alpha = dq_\alpha dp_\alpha$  is the phase space volume of the pure state  $\phi_\alpha$ . If the knowledge possessed by the observer about the coarse-grained flow of the system is not lost over time, then the probability  $P_j$  is constant in time, i.e.

$$\frac{dP_j}{dt} = 0 \quad (7)$$

In other words, it is assumed that there is no loss of probability from  $\Omega_j$  to any other coarse-grained region  $\Omega_i$ ,  $i \neq j$ . Written in terms of the probabilities  $P_\alpha$ , the condition of no loss of coarse-grained knowledge become

$$\begin{aligned} \frac{dP_j}{dt} &= \frac{d}{dt} \int_{\Omega_j} P_\alpha dV_\alpha \\ &= \int_{\Omega_j} \left( \frac{dP_\alpha}{dt} + P_\alpha \nabla \cdot \mathbf{v} \right) dV_\alpha \\ &= 0 \end{aligned} \quad (8)$$

Since this should hold independently on the size of  $\Omega_j$ , the integrand must identically vanish, i.e.

$$\frac{dP_\alpha}{dt} + P_\alpha \nabla \cdot \mathbf{v} = 0 \quad (9)$$

This is the continuity equation for probability flow within any given coarse-grained region  $\Omega_j$ . It is referred to as the Gibbs-Liouville equation for the probability distribution within  $\Omega_j$  [1][2]. Given that information is conserved within  $\Omega_j$  it is thus obtained that the probability distribution  $P_\alpha$  is conserved, i.e. if  $\nabla \cdot \mathbf{v} = 0$  then

$$\frac{dP_\alpha}{dt} = 0 \quad (10)$$

The continuity equation can be rewritten, showing that probability is locally conserved within  $\Omega_j$ . Using the total time derivative of  $P_\alpha$ , i.e.

$$\frac{dP_\alpha}{dt} = \frac{\partial P_\alpha}{\partial t} + \nabla P_\alpha \cdot \mathbf{v} \quad (11)$$

and the product rule

$$\nabla \cdot (P_\alpha \mathbf{v}) = \nabla P_\alpha \cdot \mathbf{v} + P_\alpha \nabla \cdot \mathbf{v} \quad (12)$$

the continuity equation become

$$\frac{\partial P_\alpha}{\partial t} + \nabla \cdot (P_\alpha \mathbf{v}) = 0 \quad (13)$$

The term  $\nabla \cdot (P_\alpha \mathbf{v})$  represent the difference between the probability outflow and inflow for the pure state  $\phi_\alpha$ . If there is a net probability outflow from  $\phi_\alpha$  to the rest of  $\Omega_j$ , i.e. if

$$\nabla \cdot (P_\alpha \mathbf{v}) > 0 \quad (14)$$

then the continuity equation give that the probability for  $\phi_\alpha$  decrease with time, i.e.

$$\frac{\partial P_\alpha}{\partial t} = -\nabla \cdot (P_\alpha \mathbf{v}) < 0 \quad (15)$$

If there is a net probability inflow to  $\phi_\alpha$  from the rest of  $\Omega_j$ , i.e. if

$$\nabla \cdot (P_\alpha \mathbf{v}) < 0 \quad (16)$$

then the continuity equation give that the probability for  $\phi_\alpha$  increase with time, i.e.

$$\frac{\partial P_\alpha}{\partial t} = -\nabla \cdot (P_\alpha \mathbf{v}) > 0 \quad (17)$$

In terminology borrowed from quantum mechanics, systems which evolve in such a way that the probability distribution is conserved in time and with a total probability equal to unity are said to exhibit unitary evolution. The assumption of unitary evolution for quantum systems is a key ingredient in the formulation of quantum mechanics. In classical mechanics, unitary evolution is a direct consequence of the principle of information conservation.

#### 4 Statistical equilibrium

Consider a system which has been closed for a sufficiently long period of time such that the density of pure states within  $\Omega_j$ , and hence  $M$ , do not change with time. In this situation, the probability distribution  $P_\alpha$  has no explicit dependence on time. The continuity equation is then reduced to

$$\nabla \cdot (P_\alpha \mathbf{v}) = 0 \quad (18)$$

This is the mathematical condition the system need to satisfy in order for it to be said to exist in statistical equilibrium. In other words, a system is in statistical equilibrium if there is no net probability flow on phase space.

#### 5 The ergodic theorem

The incompressibility of the Hamiltonian flow imply that the time the system spend in any single pure state, before evolving to the next single pure state, is the same for all pure states. If this was not the case, the state points on phase space would lump together which would signify a violation of the principle of conservation of information. This imply that over the course of a long period of time, the total time spent by the system in any given pure state is expected to be the same for all pure states. This expectation, which is due to a combination of the Gibbs-Liouville theorem and the

law of large numbers, is in this article interpreted to be equivalent to the ergodic theorem [4][5][6] of statistical mechanics. Let  $n_\alpha$  denote the number of times the system occupy the pure state  $\phi_\alpha$ . The total number of times,  $n$ , the system occupy the set of  $N$  pure states within  $\Omega_j$  is then

$$n = \sum_{\alpha=1}^N n_\alpha \quad (19)$$

The ergodic theorem then say that over a long period of time, such that  $n$  is large, it is expected that the system occupy all pure states within  $\Omega_j$  an equal number of times, i.e.

$$n_\alpha = n_\beta, \quad \forall \beta \neq \alpha \in [1, N] \quad (20)$$

such that

$$n = N \cdot n_\alpha \quad (21)$$

#### 6 Microcanonical probability distribution

It is now possible to define the notion of a probability  $P_\alpha$  for the pure state  $\phi_\alpha$  of a closed system from the notion of a relative frequency<sup>1</sup>,

$$P_\alpha \equiv \lim_{n \rightarrow \infty} \frac{n_\alpha}{n} = \frac{n_\alpha}{N \cdot n_\alpha} = \frac{1}{N} \quad (22)$$

Thus, all the pure states within  $\Omega_j$  are equally probable. This imply that an observer has lost all information, down to the scale of  $V_{\Omega_j}$ , about the system, since no distinctions can be made between the possible pure states within  $\Omega_j$ . This is always true for systems in statistical equilibrium. The uniform probability distribution given by equation 22 is commonly referred to as the microcanonical [1], or fundamental [7], probability distribution.

There exist also non-uniform probability distributions. The non-uniformity arise due to interactions that the system has, or have had in the not too far distant past, with an environment. In other words, the system is, or was recently, not isolated. Due to the interaction with an environment, the density of states change with time. If the interaction is uniform on phase space, the density change uniformly on phase space. However, in general, this is not the case. An interaction, characterized by a potential energy, do depend on the specific values for the generalized coordinates. In that scenario, the density of states is a local function on phase space. This has the consequence that the total time spent by

<sup>1</sup> It must be emphasized that this relative frequency is not possible to obtain from a set of repetitive experimental measurements, since the observer, being ignorant, is not able to distinguish between the set of pure states.

the system within any given region on phase space is not necessarily the same as within any other equally-sized region. In other words, the ergodic theorem appear to be violated. Thus, not only is the probability distribution non-uniform when there is a non-negligible net interaction with the environment, it can also change over time.

From the definition of probability in statistical equilibrium it is clear that the probability for any given pure state decrease as the number of pure states  $N$  increase, i.e. as the uncertainty volume increase. In non-equilibrium, where probabilities are not equal, it is the average probability which decrease as the uncertainty volume increase.

## 7 Ergodicity breaking

It should be emphasized that the apparent violation of the ergodic theorem is not of a fundamental character. It is only due to the fact that the degrees of freedom associated with the environment cannot be excluded when defining the degrees of freedom for the system. In other words, the environment should be included in the definition of the system. If that is done then there exist no environment and hence there cannot be any net transfer of energy and particles from, or to, the system. Then, this redefined system, which take into account all degrees of freedom, even those which the experimenter may think belong to an 'environment', do indeed conserve information and ergodicity is not broken. The probability distribution for the states of this redefined system is uniform, i.e. all mixed states for any given system, assuming the system has been defined such that no degrees of freedom are being forgotten, are equally probably. In most practical situations, however, there will always exist an environment to any system under study. The question is to what degree this environment interact with the system. The weaker the interaction, the weaker is the ergodicity breaking and the closer will the system come to a uniform probability distribution.

## 8 Is information conserved or lost?

At this stage, it is necessary to clarify the notions of information loss and information conservation to avoid confusion. The process of information loss, experienced by an observer, and the notion that information is conserved seem to contradict each other, making it impossible for a given system to reach statistical equilibrium over time unless it started there. The confusion arise

due to a key difference between the statistical evolution of the system, as experienced by an ignorant observer, and the deterministic evolution of the system as described by the classical laws of motion. The subtlety which must be emphasized is that the statement of classical determinism, represented mathematically by the Gibbs-Liouville theorem, is a postulate on the fundamental character of classical systems independent on whether there exist any observer or not, whereas the notion of information loss is observer dependent. The statement of information loss tries to capture the tendency of an observer to become more ignorant over time. In conclusion, statistical equilibrium is a statement on the amount of knowledge, or information, an observer of the system possess. In statistical equilibrium, the observer is unable to make any distinctions between the possible states of the system and therefore possess zero information. This do not imply that there are no fundamental distinctions between the states of the system. The observer is simply unaware of them. In fact, if the system do fundamentally conserve information, i.e. the distinctions between the possible states of the system exist for all times, the system is fundamentally never in statistical equilibrium, from the perspective of an observer whose knowledge of the degrees of freedom for the system is perfect and complete.

Thus, the fundamental question of how a system, which is initially not in statistical equilibrium, can evolve into a state of statistical equilibrium, should be modified as follows:

*How can an observer of a given classical system, which flow on phase space according to the Hamilton equations, loose information about the system over time?*

## 9 Entropy as a measure of uncertainty

A measure for the amount of ignorance possessed by the observer, i.e. the amount of uncertainty in the determination of the pure state of the system, should depend on the probability distribution  $\{P_\alpha\}$ . This measure is denoted by  $S(\{P_\alpha\})$  and referred to as the entropy of the system. To obtain a specific form for the entropy as a function of the probability distribution, it is noted that this function should satisfy the following conditions.

- i The entropy should be zero when the observer has complete knowledge about the evolution of the system. In other words, if the observer know with absolute certainty that the system occupy a specific state  $\phi_\alpha$ , such that  $P_\alpha = 1$  and  $P_\beta = 0 \forall \beta \neq \alpha$ , the entropy must vanish.

- ii The entropy should always be either zero or a positive number, i.e.  $S \geq 0$ .
- iii The entropy should take a maximum value when the observer is maximally ignorant. This happen when the system is in statistical equilibrium. When all states are equally probable, it imply that the observer possess zero partial knowledge which can be used to distinguish between some of the features of the set of states. Thus,

$$P_\alpha = \frac{1}{N} \quad \forall \alpha \in [1, N] \rightarrow S(\{P_\alpha\}) = S_{max} \quad (23)$$

- iv The entropy should, in statistical equilibrium, be a continuously increasing function of the number of states  $N$ . In other words, when  $N$  increase, the uncertainty volume  $V_{\Omega_j}$  increase continuously.
- v The entropy should satisfy the following composition law,

$$S(\{P_\alpha\} \cdot \{P_\beta\}) = S(\{P_\alpha\}) + S(\{P_\beta\}) \quad (24)$$

This composition law is understood as follows. Let  $\Omega_j$  be divided into two subregions  $\Omega_j^\alpha$  and  $\Omega_j^\beta$  such that  $V_{\Omega_j} = V_{\Omega_j^\alpha} + V_{\Omega_j^\beta}$ . The states  $\phi_\alpha, \alpha \in [1, N_\alpha]$ , belong to  $\Omega_j^\alpha$  and the states  $\phi_\beta, \beta \in [1, N_\beta]$ , belong to  $\Omega_j^\beta$ , where  $N_\alpha + N_\beta = N$ . The corresponding probability distributions,  $\{P_\alpha\}_{\alpha=1}^{N_\alpha}$  and  $\{P_\beta\}_{\beta=1}^{N_\beta}$ , satisfy  $\sum_{\alpha=1}^{N_\alpha} P_\alpha + \sum_{\beta=1}^{N_\beta} P_\beta = 1$  and, due to them being independent of each other, their product give the probability distribution associated with the region  $\Omega_j$ , i.e.  $P(\Omega_j) = \{P_\alpha\} \cdot \{P_\beta\}$ . The composition law thus state that the total uncertainty within region  $\Omega_j$  is the sum of the uncertainties associated with the subregions of  $\Omega_j$ .

Conditions (i) and (v) suggest that the entropy has a logarithmic dependence on the probability distribution. Condition (ii) suggest that it is necessary to include an additional minus sign in the definition of the entropy. This is seen from the general definition of  $P_\alpha$ , i.e.

$$\log P_\alpha = \lim_{n \rightarrow \infty} \log \left( \frac{n_\alpha}{n} \right) = \log n_\alpha - \lim_{n \rightarrow \infty} \log n < 0 \quad (25)$$

which, for a system in statistical equilibrium become

$$\log P_\alpha = \log \frac{1}{N} = \log 1 - \log N = -\log N < 0 \quad (26)$$

Since the entropy function should act as a measure for systems both in and out of statistical equilibrium, i.e. for both uniform and non-uniform probability distributions, it is required to take the statistical average of all logarithmic contributions to the entropy, i.e.

$$S(\{P_\alpha\}) \sim - \left( \frac{n_1}{n} \log P_1 + \dots + \frac{n_N}{n} \log P_N \right) \quad (27)$$

$$\sim - \sum_{\alpha=1}^N \frac{n_\alpha}{n} \log P_\alpha \quad (28)$$

$$\sim - \sum_{\alpha=1}^N P_\alpha \log P_\alpha \quad (29)$$

This entropy function then satisfy conditions (iii) and (iv). With the proportionality constant identified with the Boltzmann constant  $k_B$ , it is referred to as the Gibbs entropy [1] and is, in the information theoretic language, identical to the Shannon entropy [8][9][10].

In conclusion, the entropy of a system measure the amount of uncertainty within the system, and it is given by the Gibbs formula

$$S(\{P_\alpha\}) = -k_B \sum_{\alpha=1}^N P_\alpha \log P_\alpha \quad (30)$$

In statistical equilibrium, the Gibbs entropy reduce to the Boltzmann entropy [4][11],

$$S = k_B \log M \quad (31)$$

It is important to emphasize that entropy is not a physical quantity in the same manner as e.g. energy. It is determined by the probability distribution of the states of the system and as such it is a quantity which depend both on the specifics of the system and of the ignorance of the observer.

## 10 The second law of thermodynamics

Given that the probability distribution is conserved, i.e. that the observer do not become more ignorant over time, the entropy, which is a function of the probability distribution, is necessarily also conserved. In practical reality, however, it is most often the case that the observer loose more and more track of the flow of the system as time evolve thus becoming more ignorant over time. The reason for this is that any motion of some object, initially fairly isolated, will interact with its environment thus involving more and more degrees of freedom over time, making it increasingly difficult for the observer to avoid information loss.

Thus, for any observer, over time, the entropy tend to increase. Therefore, over time, the observer tend to loose information about the system. Eventually, the entropy has reached a maximum value, which is when the system is in statistical equilibrium. At this point, entropy will not increase and the observer has stopped losing information. The information has been completely lost. There is zero information left since the observer is unable to make any type of distinctions between the possible states of the system. The second law of thermodynamics is thus, in terms of information, stated as:

*An observer tend to loose information about any given*

*system over time until there is none left.*

It is important to emphasize that the second law is a probabilistic law. The amount of information possessed by an observer about a system is equivalently characterized by the amount of uncertainty, i.e. the entropy, which is dependent on the probability distribution as discussed earlier.

### Acknowledgement

The author would like to express gratitude to Anke Olsen for many inspiring discussions in our office at Bacchus.

### References

1. J.W. Gibbs “Elementary principles in statistical mechanics,” Charles Scribner’s sons, New York. (1902).
2. F. Bloch “Fundamentals of statistical mechanics: manuscript and notes of Felix Bloch,” Imperial College Press and World Scientific Publishing, 3rd ed. (2000).
3. A. Henriksson “On the Gibbs-Liouville theorem in classical mechanics,” Philosophy of Science Archive (PhilSci Archive) (2019). <http://philsci-archive.pitt.edu/15985/> (accessed 2019-05-08).
4. L.E. Boltzmann “Vorlesungen über Gastheorie, vol II,” Verlag von Johann Ambrosius Barth, Leipzig (1898).
5. G.D. Birkhoff “Proof of the ergodic theorem,” Proceedings of the National Academy of Sciences (U.S.), **17**, p.656-660 (1931).
6. J.v. Neumann “Physical applications of the ergodic hypothesis,” Proceedings of the National Academy of Sciences (U.S.), **18**, p.263-266 (1932).
7. A.Ya. Khintchine “Fundamental laws of probability,” Moscow State University, Moscow, 1927; 2nd ed. GTTI, Moscow (1932).
8. C.E. Shannon “A mathematical theory of communication,” The Bell System Technical Journal, **27**, p.379-423, 623-656 (1948).
9. E.T. Jaynes “Information theory and statistical mechanics,” Physical Review, **106**, No.4, p.620-630 (1957).
10. E.T. Jaynes “Information theory and statistical mechanics II,” Physical Review, **108**, No.2, p.171-190 (1957).
11. L.E. Boltzmann “Vorlesungen über Gastheorie, vol I,” Verlag von Johann Ambrosius Barth, Leipzig (1896).