JGSP **57** (2020) 1–42

ON GIBBS STATES OF MECHANICAL SYSTEMS WITH SYMMETRIES

CHARLES-MICHEL MARLE

Communicated by XXX

Abstract. The French mathematician and physicist Jean-Marie Souriau studied Gibbs states for the Hamiltonian action of a Lie group on a symplectic manifold and considered their possible applications in Physics and Cosmology. These Gibbs states are presented here with detailed proofs of all the stated results. A companion paper to appear will present examples of Gibbs states on various symplectic manifolds on which a Lie group of symmetries acts by a Hamiltonian action, including the Poincaré disk and the Poincaré half-plane.

MSC: 53D05, 53D20, 53D17, 82B03, 82B30.
Keywords: Symplectic and Poisson manifolds, Liouville measure, Hamiltonian systems, moment maps, Gibbs states, thermodynamic equilibrium.

In memory of the French mathematician and physicist Jean-Marie Souriau (1922–2012)

1. Introduction

The French mathematician and physicist Jean-Marie Souriau (1922–2012) considered, first in [26], then in his book [27], Gibbs states on a symplectic manifold built with the moment map of the Hamiltonian action of a Lie group, and the associated thermodynamic functions. In several later papers [28–30], he developed these concepts and considered their possible applications in Physics and in Cosmology. A partial translation in English of these papers, made by Frédéric Barbaresco, is available [7].

Recently, under the name *Souriau's Lie groups thermodynamics*, these Gibbs states and the associated thermodynamic functions were considered by several scientists, notably by Frédéric Barbaresco, for their possible applications in today very fashionable scientific topics, such as Geometric Information Theory, Deep Learning and Machine Learning [3–6, 8, 23, 24]. Although including these topics in a reasearch program seems to be, nowadays, a good way to obtain a public funding, I am not going to speak about them, since they are far from my field of knowledge. I will rather stay on Gibbs states and their possible applications in classical and relativistic Mechanics.

Long before the works of Souriau, the American scientist Josiah Willard Gibbs (1839–1903) considered Gibbs states associated to the Hamiltonian action of a Lie group. In his book [12] published in 1902, he clearly described Gibbs states in which the components of the total angular momentum (which are the components of the moment map of the action of the group of rotations on the phase space of the considered system) appear, on the same footing as the Hamiltonian. He even considered Gibbs states involving conserved quantities more general than those associated with the Hamiltonian action of a Lie group. In this domain, Souriau's main merits do not lie, in my opinion, in the consideration of Gibbs states for the Hamiltonian action of a Lie group, but rather in the use of the *manifold of motions* of a Hamiltonian system instead of the use of its *phase space*, and his introduction, under the name of *Maxwell's principle*, of the idea that a symplectic structure should exist on the manifold of motions of systems encountered as well in classical Mechanics as in relativistic Physics. He therefore considered Gibbs states for Hamiltonian actions, on a symplectic manifold, of various Lie groups, including the Poincaré group, often considered in Physics as a group of symmetries for isolated relativistic systems. He was well aware of the fact that Gibbs states for the Hamiltonian action of the full considered groups may not exist, which led him to carefully discuss the physical meaning and the possible applications of Gibbs states associated to the action of some of their subgroups.

The present paper is the first of a series of two papers devoted to Gibbs states for Hamiltonian actions of Lie groups. Section 2 begins with a short history of the birth of statistical Mechanics, fommowed by a reminder about the use of Hamiltonian vector fields in Mechanics and about the Liouville measure on a symplectic manifold. The concept of *statistical state* is introduced and its physical meaning is discussed. The *entropy* of a continuous statistical state is discussed and related to Shannon's entropy of a discrete random variable used in information theory. Gibbs states in the special case in which the only conserved quantity considered is the Hamiltonian, and the associated thermodynamic functions, are then briefly presented. Their physical interpretation as states of thermodynamic equilibrium is discussed. The relation of the real parameter β used to index statistical states with the temperature is explained. Section 3 begins with the notion of *manifold of motions* of a Hamiltonian dynamical system. Then Gibbs states for the Hamiltonian action of a Lie group on a symplectic manifold are discussed, with full proofs of all the stated results. Most of these proofs can be found in Souriau's book [27], which some readers may find difficult to access. A good English translation of this book is available, which faithfully preserves the language and the notations of the author. We have chosen here to use a language and notations closer to those today used in differential geometry.

The companion paper to appear will present several examples of Gibbs states on various symplectic manifolds on which acts a Lie group, including the Poincaré disk and the Poincaré half-plane.

2. Some concepts used in statistical mechanics

2.1. The birth of statistical mechanics

In his book *Hydrodynamica* published in 1738, Daniel Bernoulli (1700–1782) considered fluids (gases as well as liquids) as made of a very large number of moving particles. He explained that the pressure in the fluid is the result of collisions of the moving particles against the walls of the vessel in which it is contained, or against the probe which measures the pressure.

Daniel Bernoulli's idea remained ignored by most scientists for more than one hundred years. It is only in the second half of the XIX-th century that some scientists, notably Rudolf Clausius (1822–1888), James Clerk Maxwell (1831–1879) and Ludwig Eduardo Boltzmann(1844–1906), considered Bernoulli's idea as reasonable. As soon as 1857, Clausius began the elaboration of a *kinetic theory of gases* aiming at the explanation of macroscopic properties of gases (such as temperature, pressure and other thermodynamic properties), starting from the equations which govern the motions of the moving particles. Around 1860, Maxwell determined the probability distribution of the moving particles velocities in a gas in thermodynamic equilibrium. For a gas not in thermodynamic equilibrium, an evolution equation for the probability distribution of kinematical states of the moving particles in phase space was obtained by Boltzmann in 1872. He obtained this now famous equation, today called the *Boltzmann equation*, by using probabilistic arguments about the way in which collisions betweeen particles can occur. He introduced a quantity, denoted by the letter H[1] which, as a functon of time, always montonically decreases. Boltzmann's H function is now identified with the opposite of the *entropy* of the gas. On this basis, Gibbs laid the foundations of a new branch of theoretical physics, which he called *statistical mechanics* [12].

In the first half of the XX-th century, scientists understood that the motions of molecules in a material body do not perfectly obey Newton's laws of classical mechanics, and that the laws of quantum mechanics should be used instead. The basic concepts of statistical mechanics established by Gibbs were general enough

[1] In Boltzmann's mind, this letter was probably the Greek boldface letter *Êta* rather than the Latin letter H.

to remain valid in this new framework, and to be used for liquids or solids as well as for gases.

2.2. Statistical states and entropy

In this subsection, after a reminder of some well known facts about the use of Hamiltonian vector fields in classical mechanics and about symplectic manifolds, the important concept of a *statistical state* is presented and the definition of its *entropy* is given.

2.2.1. The use of Hamiltonian vector fields in classical mechanics

Let us recall how the evolution with time of the state of a material body is mathematically described by a dynamical system, in the framework of classical mechanics. The physical time $\mathcal{T}$ is a one-dimensional real, oriented affine space, identified with $\mathbb{R}$ once a unit and an origin of time are chosen. The set of all possible kinematic states of the body is a symplectic manifold (M, ω), very often a cotangent bundle, traditionnaly called the *phase space* of the system. For an isolated system, a smooth real-valued function H, defined on M, called a *Hamiltonian* for the system, determines all its possible evolutions with time. Let indeed X_H be the unique smooth vector field, defined on M, which satisfies the equality

$$\mathrm{i}(X_H)\omega = -\mathrm{d}H \,. \tag{$*$}$$

It is called the *Hamiltonian vector field* admitting the function H as a Hamiltonian. Any possible evolution with time of the system is described by a smooth curve $t \mapsto \varphi(t)$, defined on an open interval in $\mathbb{R}$, with values in M, which is a maximal integral curve of the differential equation, called *Hamilton's equation*, in honour of the Irish mathematician William Rowan Hamilton (1805–1865),

$$\frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} = X_H\bigl(\varphi(t)\bigr) \,. \tag{$**$}$$

The Hamiltonian H is a *first integral* of this differential equation : it means that for each smooth curve $t \mapsto \varphi(t)$, solution of this differential equation, $H\bigl(\varphi(t)\bigr)$ is a constant.

More generally, when the system is not isolated, its Hamiltonian H is a smooth function defined on $\mathbb{R} \times M$ (or on an open subset of $\mathbb{R} \times M$) since it may depend on time. The Hamiltonian vector field X_H which admits such a function as Hamiltonian is still, for each time $t \in \mathbb{R}$, determined by equation $(*)$ above, in the righ-hand side of which the differential $\mathrm{d}H$ must be calculated, for each $t \in \mathbb{R}$, as its *partial differential* with respect to the variable $x \in M$, the time $t \in \mathbb{R}$ being considered as fixed. Therefore X_H is a *time-dependent vector field* on M, *i.e.*, a smooth map, defined on some open subset of $\mathbb{R} \times M$, with values in the tangent bundle TM, such that for each fixed $t \in \mathbb{R}$, the map $x \mapsto X_H(t, x)$ is an usual smooth vector field defined on some open subset of M. Any possible evolution with time of the system is still described by a smooth curve $t \mapsto \varphi(t)$, which is a maximal integral curve of the differential equation $(**)$ above, wich now must be written as

$$\frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} = X_H\bigl(t, \varphi(t)\bigr)\,, \qquad (***)$$

in order to indicate that X_H may depend on $t \in \mathbb{R}$ as well as on $\varphi(t) \in M$. In this case the Hamiltonian H is no more a first integral of this differential equation.

2.2.2. The Liouville measure on a symplectic manifold

Let (M, ω) be a $2n$-dimensional symplectic manifold. Let (U, φ) be an admissible chart of M. For each $x \in M$, we set

$$\varphi(x) = (q^1, \dots, q^n, p_1, \dots, p_n) \in \varphi(U) \subset \mathbb{R}^{2n}\,.$$

The chart (U, φ) is said to be *canonical*, or to be a *Darboux chart*, if the local expression of ω in U is

$$\omega = \sum_{i=1}^{n} \mathrm{d}p_i \wedge \mathrm{d}q^i\,.$$

The local coordinates $q^1, \dots, q^n, p_1, \dots, p_n$ in this chart are called *canonical coordinates* or *Darboux coordinates*. The famous *Darboux theorem*, so named in honour of the French mathematician Gaston Darboux (1842–1917), asserts that any point in M is an element of the domain of a canonical chart. By using this theorem, one can prove the existence of a unique positive measure on the Borel σ-algebra[2] of M, called the *Liouville measure*, in hounour of the French mathematician Joseph Liouville (1809–1882) and denoted by λ_ω, such that for any measurable subset A of M contained in the domain U of a canonical chart (U, φ) of M, such that $\varphi(A)$ is a bounded subset of $\mathbb{R}^{2n}$,

$$\lambda_\omega(A) = \int_{\varphi(A)} \mathrm{d}q^1 \dots \mathrm{d}q^n \mathrm{d}p_1 \dots \mathrm{d}p_n\,.$$

The Liouville measure is invariant by symplectomorphisms, which means that its direct image $\Phi_* \lambda_\omega$ by any symplectomorphism $\Phi : M \to M$ is equal to λ_ω.

[2] The Borel σ-algebra of a topological space M is the smallest family of subsets of M which contains all open subsets and is stable by complementation and by intersections of countable subfamilies. It is so named in honour of the French mathematician Émile Borel (1871–1956).

2.2.3. Definitions

Let (M, ω) be a symplectic manifold and λ_ω its Liouville measure.

1. A *statistical state* on M is a probability measure μ on the Borel σ-algebra of M. The statistical state μ is said to be *continuous* (respectively, *smooth*) when it can be written as $\mu = \rho\lambda_\omega$, where ρ is a continuous function (respectively, a smooth function) defined on M. The function ρ is then said to be the *probability density* (or simply the *density*) of the statistical state μ with respect to λ_ω.

2. Let μ be a statistical state on M and f be a function, defined on M, which takes its values in $\mathbb{R}$ or in a finite-dimensional vector space. When f is integrable on M with respect to the measure μ, its integral is called the *mean value* of f in the statistical state μ, and denoted by $\mathcal{E}_\mu(f)$. When the statistical state μ is continuous, with the continuous function ρ as probability density with respect to λ_ω, the mean value of f in the statistical state μ is, by a slight abuse of notations, denoted by $\mathcal{E}_\rho(f)$. Its expression is

$$\mathcal{E}_\rho(f) = \int_M f(x)\rho(x)\,\lambda_\omega(\mathrm{d}x)\,.$$

2.2.4. Comments about the use of statistical states

When the considered dynamical system, determined by the Hamiltonian vector field X_H, is made of a large number N of moving particles, the dimension of the symplectic manifold (M, ω) which represents the set of all its possible kinematic states is very large : at least $6N$, and even more when the particles are not treated as material points. A perfect knowledge of each element of M is not possible, which explains the use of statistical states in classical Mechanics. In this framework, when the state of the considered system at a given time t_0 is mathematically described by a statistical state μ, it means that instead of looking at the evolution in time of a unique system whose kinematical state at time t_0 is a given element $x_0 \in M$, one is going to look at the evolution in time of a whole family of similar systems. The evolution with time of each of these systems is described by the differential equation determined by X_H, and its kinematical states at time t_0 can be any point in the support[3] of μ.

When, instead of classical mechanics, quantum mechanics is used for the mathematical description of the evolution with time of the state of a physical system,

[3] The support of a measure μ defined on the Borel σ-algebra of a topological state M is the closed subset of M, complementary to the open subset made by points contained in an open subset U of M such that $\mu(U) = 0$.

the use of statistical states is not due to an imperfect knowledge of the initial state of the system : it is mandatory. Informations about the evolution with time of the state of a system given by quantum mechanics are indeed always probabilistic. By nature, quantum mechanics is always statistical.

2.2.5. Examples

1. Let $x_1\,,\,x_2\,,\ldots,\,x_N$ be N pairwise distinct points in M, and $k_1\,,\,k_2\,,\ldots,\,k_N$ be real numbers satisfying $k_i>0$ for all $i\in\{1,\ldots,N\}$ and $\sum_{i=1}^N k_i=1$. For each $i\in\{1,\ldots,N\}$, let δ_{x_i} be the Dirac measure at x_i, whose value $\delta_{x_i}(A)$ for a measurable subset A of M is 0 when $x_i\notin A$ and 1 when $x_i\in A$. The measure $\mu=\sum_{i=1}^N k_i\delta_{x_i}$ is a statistical state, which is neither continuous, nor smooth. The mean value of a function f in the statistical state μ is $\sum_{i=1}^N k_i f(x_i)$.

For each $i\in\{1,\ldots,N\}$, the measure δ_{x_i} is a statistical state in which the kinematical state of the system is the point x_i, with a probability 1. One can say that δ_{x_i} is a state in the usual sense. In the statitical state μ, the kinematical state of the system is a random variable which can take each value x_i with the probability k_i.

2. Still under the same assumptions, for each $i\in\{1,\ldots,N\}$, let U_i be a neighbourhood of x_i and φ_i be a positive valued, smooth function, with compact support contained in U_i, satisfying the equality $\int_M f(x)\lambda_\omega(\mathrm{d}x)=1$. The measure ν whose probability density with respect to the Liouville measure λ_ω is $\rho_\nu=\sum_{i=1}^N k_i\varphi_i$ is a smooth statistical state, which can be considered as a smooth approximation of the discrete statistical state μ considered above. Such smooth approximations of non-smooth statistical states were extensively used by the founder of geostatistics, the French mathematician and geologist Georges Matheron (1930–2000) [22] .

2.2.6. Remark

Let μ be a continuous statistical state on the symplectic manifold (M,ω) and ρ its probability density with respect to the Liouville measure λ_ω. For each measurable subset A of M, we have

$$\mu(A)=\int_A\rho(x)\,\lambda_\omega(\mathrm{d}x)\,,\quad \text{so for } A=M,\quad \mu(M)=\int_M\rho(x)\,\lambda_\omega(\mathrm{d}x)=1\,.$$

The function ρ therefore takes its values in $\mathbb{R}^+$ and is integrable on M with respect to the Liouville measure.

2.2.7. Evolution with time of a statistical state

Let (M,ω) be a symplectic manifold, $H \in C^\infty(M,\mathbb{R})$ be a smooth Hamiltonian on M which does not depend on time and X_H be the associated Hamiltonian vector field on M. We denote by Φ^{X_H} the reduced flow[4] of X_H. If, at a time t_0, the state of the dynamical system described by X_H is a perfectly defined point $x_0 \in M$, the state of the system, at any other time t_1 at which it exists, is the point $x_1 = \Phi^{X_H}_{t_1-t_0}(x_0)$.

Let us assume that $\mu(t_0)$ is the statistical state of such a system at a given time t_0. We assume, for simplicity, that $\mu(t_0)$ is smooth and we denote by $\rho(t_0)$ its probability density with respect to the Liouville measure λ_ω. Let t_1 be another time at which the considered system still exists. The reduced flow Φ^{X_H} of the Hamiltonian vector field X_H is such that $\Phi^{X_H}_{t_1-t_0}$ is a symplectic diffeomorphism of an open subset of M onto another open subset of this manifold, whose inverse is $\Phi^{X_H}_{t_0-t_1}$. The statistical state of the system at time t_1 is therefore smooth, with a probability density $\rho(t_1)$ with respect to λ_ω, related to $\rho(t_0)$ by the equation

$$\rho(t_1) = \rho\bigl(t_0\bigr) \circ \Phi^{X_H}_{t_0-t_1} \,.$$

In other words, for any $x \in M$,

$$\rho(t_1,x) = \rho\bigl(t_0, \Phi^{X_H}_{t_0-t_1}(x)\bigr) \,.$$

2.2.8. Definition

Let ρ be the probability density, with respect to the Liouville measure λ_ω, of a continuous statistical state on the symplectic manifold (M,ω). The *entropy* of this statistical state, denoted by $s(\rho)$, is defined as follows. With the convention that when $x \in M$ is such that $\rho(x) = 0$, we set $\log\left(\dfrac{1}{\rho(x)}\right)\rho(x) = 0$, we can consider $x \mapsto \log\left(\dfrac{1}{\rho(x)}\right)\rho(x)$ as a continuous function well defined on M, taking its

[4] The *full flow*, or in short the *flow*, of a smooth vector field X, which may depend on time, defined on $\mathbb{R}\times M$ (or on an open subset of $\mathbb{R}\times M$) is the map Ψ^X, defined on an open subset of $\mathbb{R}\times\mathbb{R}\times M$, taking its values in M, such that for each $t_0 \in \mathbb{R}$ and each $x_0 \in M$, the maximal solution φ of the differential equation determined by X which satisfies $\varphi(t_0) = x_0$ is the map $t \mapsto \Psi^X(t,t_0,x_0)$. When X does not depend on time, $\Psi^X(t,t_0,x_0)$ only depends on $t-t_0$ and x_0. So instead of the full flow Ψ^X, one can use the *reduced flow* Φ^X, defined on an open subset of $\mathbb{R}\times M$ by the equality $\Phi^X(t,x_0) = \Psi^X(t,0,x_0)$. One often write $\Phi^X_t(x_0)$ to emphasize the fact that Φ^X_t is a diffeomorphism between two open subsets of M.

values in $\mathbb{R}$. When this function is integrable on M with respect to the Liouville measure λ_ω, we set

$$s(\rho) = \int_M \log\left(\frac{1}{\rho(x)}\right)\rho(x)\lambda_\omega(\mathrm{d}x) = -\int_M \log\bigl(\rho(x)\bigr)\rho(x)\lambda_\omega(\mathrm{d}x)\,.$$

Otherwise, we set

$$s(\rho) = -\infty\,.$$

The map $\rho \mapsto s(\rho)$ so defined on the set of all continuous probability densities on M is called the *entropy functional*.

2.2.9. Comments about entropy

1. The concept of *entropy* is due to Rudolf Clausius, who used it to formulate precisely the second principle of thermodynamics.

2 The entropy of a real system in Physics is always positive. The third law of thermodynamics states that the entropy of a system in thermodynamic equilibrium, when its state of minimal energy is unique, decreases towards 0 when its absolute temperature decreases towards 0 degree Kelvin. Physicists therefore consider as an unacceptable anomaly the fact that the entropy functional can take negative values, and are scandalized at the sight of $-\infty$ as a possible value of entropy. Indeed, such a value is in clear conflict with Heisenberg's principle of uncertainty. The von Neumann entropy[5], used in quantum statistical mechanics, is always positive, and the entropy defined in 2.2.8 is only its imperfect classical approximation.

3. In his famous paper [25], written during the second world war and published in 1948, the American mathematician, electrical engineer and cryptographer Claude Elwood Shannon (1916–2001) laid the foundations of information theory. He defined in this paper a concept of *entropy* whose *opposite* can be used as measurement of the information contained in a message, and considered its evolution when the message is transmitted through a telecommunications channel. Curiously enough, by reference to Boltzmann's works, the notation he used for his entropy is the letter H, although he observed that his entropy's expression is similar to the expression of the *opposite* of Boltzmann's H-function. For a random variable X which can take N possible values x_i, repectively with the probabilities k_i[6] $(1 \leq i \leq n)$, the k_i

[5] In quantum statistical mechanics, the von Neumann entropy of a state mathematically described by a *density matrix* ρ is the trace of $-\rho \ln \rho$. It was defined and extensively used by the Hungarian-American universal scientist John von Neumann (1903–1957).

[6] The notation used by Shannon for the probability of x_i is p_i, $1 \leq i \leq N$. Here I use k_i instead to avoid any risk of confusion with the Darboux coordinates p_i in a canonical chart of a symplectic manifold.

satisfying $k_i \geq 0$ and $\sum_{i=1}^{N} k_i = 1$, Shannon defined its entropy $H(X)$ by stating

$$H(X) = \sum_{i=1}^{N} \log\left(\frac{1}{k_i}\right) k_i = -\sum_{i=1}^{N} (\log k_i) k_i \,,$$

with the usual convention $0 \log 0 = 0$. In Appendix 2 of his above cited paper, page 49, he proved that up to multiplication by a strictly positive constant, his entropy is the only function which satisfies the following three very reasonable requirements.

- The function H must continuously depend on the probabilities k_i, $1 \leq i \leq N$.
- When the k_i are all equal to $1/N$ the function $N \mapsto H(1/N, \ldots, 1/N)$ (N terms) must increase monotonically with N.
- When some possible values of the random variable X are obtained as the result of two successive choices, the value of $H(X)$ must be equal to the weighted sum of the individual values of H. For example, for a random variable X with the three possible values : x_1 with probability $k_1 = 1/2$, x_2 with probability $k_2 = 1/3$ and x_3 with probability $k_3 = 1/6$, the values x_1, x_2 and x_3 can be obtained in two steps. In the first step, a first trial is done in which one looks at the value taken by a random variable Y with two possible values, y_1 and y_2, both obtained with probability $1/2$. In the second step, if the value taken by Y is y_1, one states that the value taken by X is x_1; if the value taken by Y is y_2, one looks at the value taken by a random variable Z with two possible values, z_1 with probability $2/3$ and z_2 with probability $1/3$. If the value taken by Z is z_1, one states that the value taken by X is x_2, and if the value taken by Z is z_2, one states that the value taken by X is x_3. The equality that the function H is required to satisfy is

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H\left(\frac{2}{3}, \frac{1}{3}\right) \,.$$

Interested readers are referred to Alain Chenciner's paper [10] for a more detailed account of Claude Shannon's works and their influence on today's science.

4. The American physicist Edwin Thompson Jaynes (1922–1998) observed, in [15, 16] (see also [31]), that the definition 2.2.8 of entropy for a continuous statistical state of probability density ρ with respect to the Liouville measure,

$$s(\rho) = \int_M \log\left(\frac{1}{\rho(x)}\right) \rho(x) \lambda_\omega(\mathrm{d}x) = -\int_M \log\bigl(\rho(x)\bigr) \rho(x) \lambda_\omega(\mathrm{d}x) \,,$$

is not a correct adaptation of Shannon's entropy for a discrete statistical state which can take N distinct values x_i, with the respective probabilities k_i,

$$H(X) = -\sum_{i=1}^{N} (\log k_i) k_i\,, \quad \text{with } k_i \geq 0 \text{ for all } i \in \{1, \dots, N\} \text{ and } \sum_{i=1}^{N} k_i = 1\,.$$

While $H(X)$ is always a dimensionless number satisfying $H(X) \geq 0$, and $H(X) = 0$ if and only if there exists only one integer $i \in \{1, \dots, N\}$ such that $k_i = 1$, all other k_j, fo $j \neq i$, being equal to 0, the above expression of $s(\rho)$ depends on the chosen units. Indeed in this expression, while $\rho(x)\lambda_\omega(\mathrm{d}x)$ is dimensionless, $\rho(x)$, as well as $\lambda_\omega(\mathrm{d}x)$ are not dimensionless. A change of the units (of length, time and mass) changes the value of the the term $\log\bigl(\rho(x)\bigr)$ by addition of a constant, which can be either positive or negative. Therefore when one uses definition 2.2.8, the sign of entropy does not have any physical meaning. In the above cited papers of Jaynes, the author considered problems in statistics more general than those encountered in classical statistical mechanics, in which the Liouville measure may not be available. He proposed to replace, in the expression of the entropy $s(\rho)$, the term $\log\left(\dfrac{1}{\rho(x)}\right)$ by $\log\left(\dfrac{m(x)}{\rho(x)}\right)$, where $m(x)$ is the probability density of a reference statistical state with respect to which the entropy $s(\rho)$ is evaluated. Of course the probability densities $m(x)$ and $\rho(x)$ must be taken with respect to the same measure. In the framework of classical statistical mechanics, this measure is the Liouville measure λ_ω, so the correction proposed by Jaynes can be written

$$s_{\text{Jaynes}}(\rho) = \int_M \log\left(\frac{m(x)}{\rho(x)}\right) \rho(x)\lambda_\omega(\mathrm{d}x)\,.$$

Probably because he considered problems in which the Liouville measure did not appear, Jaynes did not clearly state how $m(x)$ should be chosen, although he recommanded the use of a probability density invariant by the group of automorphisms of the considered measurable space. Therefore it seems that in the framework of classical statistical mechanics, when the support W of ρ[7] is of finite λ_ω-measure, one should use the following probability density :

$$m(x) = \begin{cases} \dfrac{1}{\lambda_\omega(W)} & \text{when } x \in W, \\ 0 & \text{when } x \notin W. \end{cases}$$

With this choice of m, $s(\rho)$ and $s_{\text{Jaynes}}(\rho)$ are related by

$$s_{\text{Jaynes}}(\rho) = s(\rho) - \log\bigl(\lambda_\omega(W)\bigr)\,.$$

[7] The support of ρ is the closure of the subset of M made of points $x \in M$ such that $\rho(x) \neq 0$.

The corrected entropy $s_{\text{Jaynes}}(\rho)$ proposed by Jaynes is dimensionless. It differs from the entropy $s(\rho)$ of Definition 2.2.8 only by a constant, which depends on the units chosen for time, length and mass, and can take negative as well as positive values.

In calculus of variations, it may be useful to consider infinitesimal variations of ρ whose support does not always remain contained in the support of ρ. Instead of the support of ρ, one should take for W, in the above formula, an open subset of M of finite λ_ω-measure which contains the support of ρ.

5. For a better understanding of how the entropies of continuous and discrete statistical states are related, let us consider the process of discretization of a continuous statistical state. As above, we assume that the support W of the probability density ρ is of finite λ_ω-measure. For simplicity[8], we moreover assume that $\rho_{\max} = \sup_{x\in M} \rho(x)$ too is finite and that, for each real r satisfying $0 \leq r \leq \rho_{\max}$,

$$\lambda_\omega\Bigl(\bigl\{x \in W \bigm| \rho(x) = r\bigr\}\Bigr) = 0\,.$$

For each $r \geq 0$, let us set

$$G(r) = \lambda_\omega\Bigl(\bigl\{x \in W \bigm| 0 \leq \rho(x) \leq r\bigr\}\Bigr)\,.$$

Then G is a continuous and monotonically increasing function which takes all values in the closed interval $[0, \lambda_\omega(W)]$. Let N be an integer satisfying $N > 2$. There exist N real numbers r_i^N, $1 \leq i \leq N$, such that for each $i \in \{1, \dots, N\}$

$$G(r_i^N) = \frac{i\lambda_\omega(W)}{N}\,.$$

We set

$$V_1^N = \bigl\{x \in W \bigm| 0 \leq \rho(x) \leq r_1\bigr\}\,,$$

and, for each $i \in \{2, \dots, N\}$,

$$V_i^N = \bigl\{x \in W \bigm| r_{i-1} < \rho(x) \leq r_i\bigr\}\,.$$

The V_i^N are measurable subsets of M which satisfy, for $1 \leq i, j \leq N$,

$$\lambda_\omega(V_i^N) = \frac{\lambda_\omega(W)}{N}\,, \quad V_i^N \cap V_j^N = \emptyset \text{ if } i \neq j\,, \quad \bigcup_{i=1}^N V_i^N = W\,.$$

[8] These assumptions could probably be avoided with the use of more sophisticated concepts in integration theory, such as the Stieltjes integral, so named in honour of the Dutch mathematician Thomas Joannes Stieltjes (1856–1892).

Now we set, for each $i \in \{1, \dots, N\}$,

$$k_i^N = \int_{V_i^N} \rho(x)\lambda_\omega(\mathrm{d}x)\,, \quad \rho_i^N = \frac{N k_i^N}{\lambda_\omega(W)}\,.$$

We have

$$0 \leq k_i^N \leq 1 \text{ for each } i \in \{1, \dots, N\}\,, \quad \sum_{i=1}^{N} k_i^N = 1\,.$$

Let ρ^N be the function defined on M by

$$\rho^N(x) = \begin{cases} \dfrac{N k_i^N}{\lambda_\omega(W)} & \text{if } x \in V_i^N,\, 1 \leq i \leq N, \\ 0 & \text{if } x \notin \bigcup_{i=1}^N V_i^N = W\,. \end{cases}$$

The function ρ^N is everywhere ≥ 0 on M, and only takes N distinct non-zero values. It is a discrete approximation of the probability density ρ, which satisfies

$$\int_M \rho^N(x)\lambda_\omega(\mathrm{d}x) = \sum_{i=1}^{N} k_i^N = 1\,.$$

The function ρ^N is therefore the probability density of a statistical state on M. Although it is not continuous, we can use Definition 2.2.8 to calculate $s(\rho^N)$. We obtain

$$s(\rho^N) = \sum_{i=1}^{N} k_i^N(-\log k_i^N) + \log\bigl(\lambda_\omega(W)\bigr) - \log N\,.$$

We observe that the term $\sum_{i=1}^N k_i(-\log k_i)$ is the Shannon entropy $H(X^N)$ of a random variable X^N wich can take N distinct values, for example the values $1, \dots, N$, with the respective probabilities $k_1^N, \dots, k_N^N$. So we can write

$$H(X^N) = s(\rho^N) - \log\bigl(\lambda_\omega(W)\bigr) + \log N = s_{\mathrm{Jaynes}}(\rho^N) + \log N\,.$$

When $N \to +\infty$, $s_{\mathrm{Jaynes}}(\rho^N) \to s_{\mathrm{Jaynes}}(\rho)$ and $\log N \to +\infty$. The above equality proves that when $N \to +\infty$, the Shannon entropy of the discrete approximation, by a random variable X^N which can take N distinct non-zero values, of the continuous statistical state of probability density ρ, does not remain bouded and increases as fast as $\log N$.

6. During the years 1950–1960, several scientists, notably Edwin Thompson Jaynes cited above (see also [13, 14] by the same author) and the American mathematician George Whitelaw Mackey (1916–2006) [17], proposed the use of information theory in thermodynamics.

Interested readers are referred to Roger Balian's paper [1], in which they will find a clear account of the use of probability concepts in physics and of information theory in quantum mechanics.

2.2.10. Proposition

On a symplectic manifold (M, ω), we consider a smooth Hamiltonian $H \in C^\infty(M, \mathbb{R})$ which does not depend on time. Let X_H be the associated Hamiltonian vector field on M. Let $\rho(t_0)$ be the probability density of a smooth statistical state of the dynamical system determined by X_H at a time t_0. The probability density $\rho(t_1)$ of the statistical state of the system at any other time t_1 at which the system still exists is such that

$$s\bigl(\rho(t_1)\bigr) = s\bigl(\rho(t_0)\bigr)\,.$$

In other words, the entropy of the statistical state of the system remains constant as long as this statistical state exists.

Proof: As seen in 2.2.7, $\rho(t_1) = \rho(t_0) \circ \Phi^{X_H}_{t_0 - t_1}$, so for each $x \in M$,

$$\rho(t_0, x) = \rho\bigl(t_1, \Phi^{X_H}_{t_1 - t_0}(x)\bigr)\,.$$

When $s\bigl(\rho(t_0)\bigr) \neq -\infty$, we can write

$$\begin{aligned}
s\bigl(\rho(t_0)\bigr) &= \int_M \log\left(\frac{1}{\rho(t_0, x)}\right) \rho(t_0, x) \lambda_\omega(\mathrm{d}x) \\
&= \int_M \log\left(\frac{1}{\rho\bigl(t_1, \Phi^{X_H}_{t_1 - t_0}(x)\bigr)}\right) \rho\bigl(t_1, \Phi^{X_H}_{t_1 - t_0}(x)\bigr)(\lambda_\omega)(\mathrm{d}x) \\
&= \int_M \log\left(\frac{1}{\rho(t_1, y)}\right) \rho(t_1, y) (\Phi^{X_H}_{t_1 - t_0})_*(\lambda_\omega)(\mathrm{d}y) \\
&= \int_M \log\left(\frac{1}{\rho(t_1, y)}\right) \rho(t_1, y) \lambda_\omega(\mathrm{d}y) \\
&= s\bigl(\rho(t_1)\bigr)\,,
\end{aligned}$$

where we have used the change of integration variable $y = \Phi^{X_H}_{t_1 - t_0}(x)$ and the invariance of the Liouville measure by symplectomorphism (2.2.2). When $s\bigl(\rho(t_0)\bigr) = -\infty$, the same calculation leads to a divergent integral for the expression of $s\bigl(\rho(t_1)\bigr)$, which therefore is equal to $-\infty$. ■

2.3. Gibbs states for a Hamiltonian system

In this subsection, H is a smooth Hamiltonian which does not depend on time, defined on a symplectic manifold (M,ω), and X_H is the associated Hamiltonian vector field. The *Gibbs states* defined here are built with the Hamiltonian H as the only conserved quantity. Their main properties are briefly indicated. Gibbs state for the Hamiltonian action of a Lie group are considered in Section 3.

2.3.1. Proposition

Under the assumptions and with the notations of Subsection 2.3, let ρ be the probability density, with respect to the Liouville measure λ_ω, of a smooth statistical state on M. We assume that ρ is such that the integrals which define the entropy $s(\rho)$ (Definition 2.2.8) and the mean value $\mathcal{E}_\rho(H)$ of the Hamiltonian H (Definition 2.2.3) are convergent and can be differentiated under the sign $\int$ with respect to infinitesimal variations of ρ. The entropy function s is stationary at ρ with respect to smooth infinitesimal variations of ρ which leave fixed the mean value of H if and only if there exists a real $\beta \in \mathbb{R}$ such that, for every $x \in M$,

$$\rho(x) = \frac{1}{P(\beta)}\exp\bigl(-\beta H(x)\bigr)\,, \quad \textit{with} \quad P(\beta) = \int_M \exp\bigl(-\beta H(x)\bigr)\lambda_\omega(\mathrm{d}x)\,.$$

Proof: Let $\tau \mapsto \rho_\tau$ be a smooth infinitesimal variation of ρ which leaves fixed the mean value of H. Since $\displaystyle\int_M \rho_\tau(x)\lambda_\omega(\mathrm{d}x)$ and $\displaystyle\int_M \rho_\tau(x)H(x)\lambda_\omega(\mathrm{d}x)$ do not depend on τ, it satisfies, for all $\tau \in]-\varepsilon,\varepsilon[$,

$$\int_M \frac{\partial\rho(\tau,x)}{\partial\tau}\lambda_\omega(\mathrm{d}x) = 0\,, \int_M \frac{\partial\rho(\tau,x)}{\partial\tau}H(x)\lambda_\omega(\mathrm{d}x) = 0\,.$$

Moreover an easy calculation leads to

$$\frac{\mathrm{d}s(\rho_\tau)}{\mathrm{d}\tau}\Bigm|_{\tau=0} = -\int_M \frac{\partial\rho(\tau,x)}{\partial\tau}\Bigm|_{\tau=0}\bigl(1+\log\bigl(\rho(x)\bigr)\lambda_\omega(\mathrm{d}x)\,.$$

By a well known result in calculus of variations, this implies that the entropy functional is stationary at ρ with respect to smooth infinitesimal variations of ρ which leave fixed the mean value of H, if and only if there exist two real constants α and β, the *Lagrange multipliers*, such that, for every $x \in M$,

$$1 + \log(\rho(x)) + \alpha + \beta H(x) = 0\,,$$

which leads to

$$\rho(x) = \exp\bigl(-1-\alpha-\beta H(x)\bigr)\,.$$

By writing that $\int_M \rho(x)\lambda_\omega(\mathrm{d}x) = 1$, we see that α is determined by β:

$$\exp(1+\alpha) = P(\beta) = \int_M \exp\bigl(-\beta H(x)\bigr)\lambda_\omega(\mathrm{d}x)\,.$$

■

2.3.2. Definition

Let $\beta \in \mathbb{R}$ be a real which satisfies the conditions of Proposition 2.3.1. The smooth statistical state whose probability density, with respect to the Liouville measure λ_ω, is

$$\rho_\beta(x) = \frac{1}{P(\beta)} \exp\bigl(-\beta H(x)\bigr)\,, \quad x \in M\,,$$

with

$$P(\beta) = \int_M \exp\bigl(-\beta H(x)\bigr)\lambda_\omega(\mathrm{d}x)\,,$$

is called the *Gibbs state* associated to (or indexed by) β. The function P of the real variable β is called the *partition function* of the dynamical system determined by the Hamiltonian vector field X_H.

2.3.3. Proposition

Let $\beta \in \mathbb{R}$ be a real which satisfies the conditions of Proposition 2.3.1. The probability density ρ_β of the corresponding Gibbs state (Definition 2.3.2) remains invariant under the flow of the Hamiltonian vector field X_H.

Proof: Since the Hamiltonian H does not depend on time, it is a first integral of the differential equation determined by X_H, *i.e.*, it keeps a constant value on each integral curve of X_H. Therefore ρ_β keeps a constant value on each integral curve of X_H. ■

2.3.4. Some properties of Gibbs states

We have seen (Proposition 2.3.1) that the entropy functional s is stationary at each Gibbs state with respect to all infinitesimal variations of its probability density which leave invariant the mean value of the Hamiltonian H. A stronger result holds: given any Gibbs state of probability density ρ_β, on the set of all continuous statistical states whose probability density ρ is such that $\mathcal{E}_\rho(H) = \mathcal{E}_{\rho_\beta}(H)$, the entropy functional s reaches its only strict maximum at the Gibbs state of probability density ρ_β.

When the set Ω of reals β for which a Gibbs state indexed by β exists is not empty, this set is an open interval $\,]\,a, b\,[\,$ of $\mathbb{R}$, where either $a \in \mathbb{R}$, or $a = -\infty$, either $b \in \mathbb{R}$ and $b > a$, or $b = +\infty$. When in addition H is bounded from below, *i.e.*, when there exists $m \in \mathbb{R}$ such that, for any $x \in M$, $m \leq H(x)$, the open interval Ω is unbound on the right side, *i.e.*, $\Omega =]\,a, +\infty\,[$ where either $a \in \mathbb{R}$, or $a = -\infty$.

We have already defined on Ω the partition function P (Definition 2.3.2). Other functions can be defined on Ω as follows. For each $\beta \in \Omega$, the entropy $s(\rho_\beta)$ exists of course, and one can prove that the mean value $\mathcal{E}_{\rho_\beta}(H)$ (Definition 2.2.3) of the Hamiltonian H, in the Gibbs state indexed by β, exists too, as well as $\mathcal{E}_{\rho_\beta}(H^2)$ and $\mathcal{E}_{\rho_\beta}\Big(\big(H - \mathcal{E}_{\rho_\beta}(H)\big)^2\Big)$. So we can set

$$S(\beta) = s(\rho_\beta)\,, \quad E(\beta) = \mathcal{E}_{\rho_\beta}(H)\,, \quad \beta \in \Omega\,.$$

The functions P (partition function), E (mean value of the Hamiltonian, considered by physicists as the energy) and S (entropy) so defined are of class C^∞ on Ω and satisfy, for any $\beta \in \Omega$,

$$\begin{aligned}
P(\beta) &> 0\,,\\
E(\beta) &= -\frac{1}{P(\beta)}\frac{\mathrm{d}P(\beta)}{\mathrm{d}\beta} = \frac{\mathrm{d}\big(-\log P(\beta)\big)}{\mathrm{d}\beta}\,,\\
\frac{\mathrm{d}E(\beta)}{\mathrm{d}\beta} &= \frac{\mathrm{d}^2\big(-\log P(\beta)\big)}{\mathrm{d}\beta^2} = -\mathcal{E}_{\rho_\beta}\Big(\big(H - \mathcal{E}_{\rho_\beta}(H)\big)^2\Big)\,,
\end{aligned}$$

$$\begin{aligned}
S(\beta) &= \log P(\beta) + \beta E(\beta) = \beta\frac{\mathrm{d}\big(-\log P(\beta)\big)}{\mathrm{d}\beta} - \big(-\log P(\beta)\big)\,,\\
\frac{\mathrm{d}S(\beta)}{\mathrm{d}\beta} &= \beta\frac{\mathrm{d}E(\beta)}{\mathrm{d}\beta}\,.
\end{aligned}$$

The above expression of $\dfrac{\mathrm{d}E(\beta)}{\mathrm{d}\beta}$ proves that $\beta \mapsto E(\beta)$ is a non-increasing function. When the Hamiltonian H is not a constant, for each $\beta \in \Omega$, the continuous function defined on M $x \mapsto \big(H(x) - \mathcal{E}_{\rho_\beta}(H)\big)^2$ takes its values in $\mathbb{R}^+$ and is not always equal to 0. Its mean value $\mathcal{E}_{\rho_\beta}\Big(\big(H - \mathcal{E}_{\rho_\beta}(H)\big)^2\Big)$ is therefore > 0, which proves that $\beta \mapsto E(\beta)$ is a strictly decreasing function on Ω. The map E is open, and is a diffeomorphism of Ω onto its image Ω^*.

The above expression of $S(\beta)$ shows that the functions $\beta \mapsto -\log\bigl(P(\beta)\bigr)$ and $\beta \mapsto S(b)$ are *Legendre transforms* of each other[9]. They are indeed linked by the same relation as that which, in calculus of variations, links a hyper-regular Lagrangian with the associated energy. Here the hyper-regular "Lagrangian", defined on Ω, is $\beta \mapsto -\log P(\beta)$, the Legendre map is the diffeomorphism $E : \Omega \to \Omega^*$, the "energy", defined on Ω, is $\beta \mapsto S(\beta)$, and the "Hamiltonian", defined on Ω^*, is $S \circ E^{-1}$. By using the above expression of $\dfrac{\mathrm{d}S(\beta)}{\mathrm{d}\beta}$, we can write

$$E^{-1}(e) = \frac{\mathrm{d}\bigl(S \circ E^{-1}(e)\bigr)}{\mathrm{d}e}\,, \quad e \in \Omega^*\,.$$

As soon as 1869, the Legendre transform was used in thermodynamics by the French scientist François Massieu [2, 19–21].

The results stated here without proof are proven below, in a more general setting, for Gibbs states associated to the Hamiltonian action of a Lie group (Propositions 3.1.3 and 3.3.6, Remarks 3.2.5).

2.3.5. Gibbs states, temperatures and thermodynamic equilibria

Let us now assume that the dynamical system determined by the Hamiltonian vector field X_H mathematically describes the evolution with time of a physical system, an object of the real world. Physicists consider each Gibbs state of the considered dynamical system, indexed by some $\beta \in \mathbb{R}$, as the mathematical description of a *state of thermodynamic equilibrium* of the corresponding physical system, and β as a quantity related to the *absolute temperature* T of the physical system by the equality

$$\beta = \frac{1}{kT}\,, \tag{$*$}$$

where k is a constant which depends on the chosen units, called *Boltzmann's constant*. This identification of Gibbs states with thermodynamic equilibria is justified by the following property. Let us consider two similar physical systems, mathematically described by two Hamiltonian systems, whose Hamiltonians are, respectively, H_1 defined on the symplectic manifold (M_1, ω_1) and H_2 defined on the symplectic manifold (M_2, ω_2). We first assume that they are independent and both in a Gibbs state. We denote by ρ_{1,β_1} and ρ_{2,β_2} the probability densities of the Gibbs states, indexed by the reals β_1 and β_2, in which these two systems are, respectively. Let $E_1(\beta_1)$ and $E_2(\beta_2)$ be the corressponding mean values of their

[9] The Legendre transform is so named in honour of the French mathematician Adrien-Marie Legendre (1752–1833).

Hamiltonians. Let us now assume that the two systems are coupled in a way allowing an exchange of energy between them. For example, the corresponding objects of the real world can be two vessels containing a gas, separated by a wall allowing a transfer of heat between them. Coupled together, they make a new physical system, mathematically described by a Hamiltonian system on the symplectic manifold $(M1 \times M_2,\ \omega_{\text{new}} = p_1^*\omega_1 + p_2^*\omega_2)$, where $p_1 : M_1 \times M_2 \to M_1$ and $p_2 : M_1 \times M_2 \to M_2$ are the canonical projections. The Hamiltonian of this new system can be made as close to $H_1 \circ p_1 + H_2 \circ p_2$ as one wishes, by making very small the coupling between the two systems. We can therefore consider $H_1 \circ p_1 + H_2 \circ p_2$ as a reasonable approximation of the Hamiltonian of the new system. When the two subsystems are in the Gibbs states indexed, respectively, by β_1 and by β_2, the new system made of these two coupled subsystems is in the statistical state of probability density $\rho_{1,\beta_1} \circ p_1 + \rho_{2,\beta_2} \circ p_2$, and its entropy is $S_1(\beta_1) + S_2(\beta_2)$. If $\beta_1 \neq \beta_2$, the new system is not in a Gibbs state. Let us indeed assume, for example, that $\beta_1 < \beta_2$. If a transfer of energy between the two subsystems occurs, in which the energy of the first subsystem decreases while the energy of the second subsystem increases by an equal amount, the modified Gibbs state of the first subsystem becomes indexed by $\beta_1' > \beta_1$ and that of the second subsystem by $\beta_2' < \beta_2$ since, as seen in 2.3.4, for $i = 1$ as well as for $i = 2$, we have $\dfrac{\mathrm{d}E_i(\beta_i')}{\mathrm{d}\beta_i'} < 0$. As long as $\beta_1' < \beta_2'$, such an energy transfer between the two subsystems results in an increase of the entropy of the total new system, until $\beta_1' = \beta_2' = \beta_n$, which indexes the Gibbs state of the new system for a mean value of its Hamiltonian $E_1(\beta_1) + E_2(\beta_2)$. We have of course $\beta_1 < \beta_n < \beta_2$, which proves that when the state of the new system evolves from its initial state towards its Gibbs state, the energy flow goes from the subsystem whose Gibbs initial state is indexed by the smaller β_1 towards the subsystem whose initial Gibbs sate in indexed by the larger β_2. This result is in agreement with everydays's experience, since equality $(*)$ implies that when $\beta > 0$, a smaller value of β corresponds to a higher temperature.

2.3.6. Evolution towards a thermodynamic equilibrium

In the real world, the state of an approximately isolated system often evolves with time towards a state of thermodynamic equilibrium. When such a state is approximately reached, it remains approximately stationary, with small fluctuations. Let us mathematically modelize this evolution as the variation with time of the statistical state of a Hamiltonian dynamical system, whose smooth Hamiltonian H is defined on a very high-dimensional symplectic manifold (M, ω). Proposition 2.3.3 above, which states that Gibbs states do not change with time, seems to be in reasonably good agreement with the identification of Gibbs states with thermodynamic equilibria, although it does not explain fluctuations which are experimentally obseved in thermodynamic equilibria. On the contrary, proposition 2.2.10 above, which states that as long as any smooth statistical state exists, its entropy remains constant, is in clear disagreement with the behaviour of isolated systems in the real world. The mathematical description of the evolution with time of a real physical system by the dynamical system determined by the Hamiltonian vector field of a smooth Hamiltonian which does not depend on time, together with Definition 2.2.8 of the entropy, can be used only for reversible systems. It cannot be used to describe the evolution with time of some statistical states towards the corresponding Gibbs state.

3. Gibbs states for Hamiltonian actions of Lie groups

The general definition of a Gibbs states is very natural : it amounts to introduce, in the definition of a statistical state, not only the Hamiltonian, but other conserved quantities too, on the same footing as the Hamiltonian. One may even forget the Hamiltonian and consider only the moment map of the Hamiltonian action.

This idea is already present in the book published by Gibbs in 1902 ([12], chapter I, page 42 and the following pages), the conserved quantities other than the Hamiltonian being the components of the total angular momentum.

Following an idea first proposed around 1809 by Joseph Louis Lagrange (1736–1813), Jean-Marie Souriau defined statistical states on the *manifold of motions* of a Hamiltonian dynamical system, instead of on its *phase space*. This approach allows a more natural treatment on the same footing of both the Hamiltonian and other conserved quantities, because the action of the group of translations in time, which may act only locally on the phase space, always acts globally on the space of motions. The concept of manifold of motions and its properties are presented in Subsection 3.1 below. Gibbs states for a Hamiltonian action of a Lie group on a symplectic manifold are then defined (Definition 3.1.4), together with generalized temperatures and partition functions, and their main property (maximality of entropy) is proven (Proposition 3.1.6). Thermodynamic functions associated to a Gibbs state (mean value of the moment map and entropy) are maps defined on the set of generalized temperatures (Subsection 3.2). The expressions of their differentials lead to the definition, on the set of generalized temperatures, of a remarkable Riemannian metric, linked to the Fisher-Rao metric of statisticians. The adjoint action on the set of generalized temperatures is considered in Subsection 3.3, in which the Riemannian metric induced on each adjoint orbit is expressed in terms of a symplectic cocycle.

3.1. Symmetries and statistical states

In this subsection we consider the dynamical system determined on a symplectic manifold (M, ω), as explained in 2.2.1, by a Hamiltonian vector field X_H whose smooth Hamiltonian H, defined on $\mathbb{R} \times M$ or on one of its open subsets, may depend on time.

3.1.1. The manifold of motions of a Hamiltonian system

Jean-Marie Souriau called *motion* of the dynamical system determined by a Hamiltonian vector field X_H any maximal solution $\varphi : t \mapsto \varphi(t)$ of Hamilton's differential equation

$$\frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} = X_H\bigl(t, \varphi(t)\bigr)\,.$$

The *manifold of motions* of the system, denoted by $\mathrm{Mot}(X_H)$, is simply the set of all motions, *i.e.*, the set of all maximal solutions φ of the above differential equation. It always has the structure of a smooth symplectic manifold. For each $t_0 \in \mathbb{R}$, the map h_{t_0} which associates, to each motion $\varphi : t \mapsto \varphi(t)$ whose interval of definition contains t_0, the point $h_{t_0}(\varphi) = \varphi(t_0) \in M$, is indeed, when the subset made of motions defined on an interval of $\mathbb{R}$ which contains t_0 is not empty, a bijection of this subset of $\mathrm{Mot}(X_H)$ onto an open subset of M. This simple fact allows the definition of a topology and a structure of smooth manifold on $\mathrm{Mot}(X_H)$ such that, for each $t_0 \in \mathbb{R}$, $h_{t_0} : \varphi \mapsto h_{t_0}(\varphi) = \varphi(t_0)$ is a diffeomorphism of the open subset of $\mathrm{Mot}(X_H)$ made of motions defined on an interval of $\mathbb{R}$ which contains t_0, onto an open subset of M. Since the reduced flow of X_H is made of symplectomorphisms, the pull-back $h_{t_0}^*\omega$ of the symplectic form ω does not depend on t_0, therefore determines globally a symplectic form $\omega_{\mathrm{Mot}(X_H)}$ on the manifold of motions $\mathrm{Mot}(X_H)$.

The manifold $\mathrm{Mot}(X_H)$ may be a *non-Hausdorff*[10] manifold, although any of its elements has an open Hausdorff neighbourhood symplectomorphic to an open subset of M.

[10]We recall that a topological space is said to be *Hausdorff* when for each pair of distinct elements x and y of this space, there exist neighbourhoods U of x and V of y such that $U \cap V = \emptyset$. This property is so named in honour of the German mathematician Felix Hausdorff (1848–1942), an important founder of topology and set theory, who after losing his Professor position at the university of Bonn, was driven to suicide by the Nazi regime.

We assume now, until the end of this section, that the Hamiltonian H does not depend on time. For a given motion $\varphi \in \mathrm{Mot}(X_H)$, the value of $H\bigl(\varphi(t_0)\bigr)$ does not depend on the choice of t_0 in the interval on which φ is defined. Therefore there exists a real-valued, smooth function H_{Mot}, defined on $\mathrm{Mot}(X_H)$, such that, for each $\varphi \in \mathrm{Mot}(X_H)$ and any t_0 in the interval on which φ is defined, $H_{\mathrm{Mot}}(\varphi) = H\bigl(\varphi(t_0)\bigr)$. Since, for each $t_0 \in \mathbb{R}$, the function H_{Mot}, restricted to the open subset made of motions whose interval of definition contains t_0, is the pull-back $h_{t_0}^*(H) = H \circ h_{t_0}$ of the Hamiltonian H, the Hamiltonian vector field $X_{H_{\mathrm{Mot}}}$ on $\mathrm{Mot}(X_H)$, restricted to this open substet of $\mathrm{Mot}(X_H)$, is the inverse image $h_{t_0}^*(X_H)$ of the Hamiltonian vector field X_H defined on M.

For any $s \in \mathbb{R}$ and any motion $\varphi :]\,a, b\,[\to M \in \mathrm{Mot}(X_H)$, defined on the interval $]\,a, b\,[\subset \mathbb{R}$, let $\Phi_{\mathrm{Mot}}(s, \varphi)$ be the parametrized curve, defined on the interval $]\,a-s, b-s\,[\subset \mathbb{R}$, with values in M,

$$\Phi_{\mathrm{Mot}}(s, \varphi)(t) = \varphi(t+s)\,, \quad t \in]\,a-s, b-s\,[\,.$$

One can easily see that $\Phi_{\mathrm{Mot}}(s, \varphi) \in \mathrm{Mot}(X_H)$ and that the map

$$\Phi_{\mathrm{Mot}} : \mathbb{R} \times \mathrm{Mot}(X_H) \to \mathrm{Mot}(X_H)$$

is a smooth action on the left of the additive Lie group $\mathbb{R}$ on the manifold of motions $\mathrm{Mot}(X_H)$. The infinitesimal generator of this action is the vector field on $\mathrm{Mot}(X_H)$, temporarily denoted by Z, defined by the equality

$$Z(\varphi) = \frac{\mathrm{d}\Phi_{\mathrm{Mot}}(s, \varphi)}{\mathrm{d}s}\bigg|_{s=0}\,, \quad \varphi \in \mathrm{Mot}(X_H)\,.$$

For any real t_0 which belongs to the open interval on which φ is defined, we have

$$Th_{t_0}\bigl(Z(\varphi)\bigr) = \frac{\mathrm{d}\Bigl(h_{t_0}\bigl(\Phi_{\mathrm{Mot}}(s, \varphi)\bigr)\Bigr)}{\mathrm{d}s}\Bigg|_{s=0} = \frac{\mathrm{d}\varphi(t_0+s)}{\mathrm{d}s}\bigg|_{s=0} = X_H\bigl(\varphi(t_0)\bigr)\,.$$

This result proves that the infinitesimal generator Z of the action Φ_{Mot} is the Hamiltonian vector field $X_{H_{\mathrm{Mot}}}$. Being generated by the flow of a Hamiltonian vector field, the action φ is therefore Hamiltonian. It admits the Hamiltonian H_{Mot} as a moment map (with the usual convention in which the Lie algebra of the additive Lie group $\mathbb{R}$ is identified with $\mathbb{R}$ with the zero bracket and its dual is too identified with $\mathbb{R}$, the pairing by duality being the usual product of reals).

The reader will observe that while the flow of the vector field X_H does not always determine a Hamiltonian action of $\mathbb{R}$ on the symplectic manifold (M, ω), but only a local Hamiltonian action, except when all the motions are defined for all $t \in \mathbb{R}$, the flow of $X_{H_{\mathrm{Mot}}}$ always determines a Hamiltonian action of $\mathbb{R}$ on $(\mathrm{Mot}(X_H), \omega_{\mathrm{Mot}(X_H)})$. However, the price paid for obtaining better properties of the flow of a Hamiltonian vector field is the fact that $\mathrm{Mot}(X_H)$ can be a non-Hausdorff manifold. For this reason, some important results, for example the theorem which asserts the unicity, for a given initial condition, of a maximal solution of a smooth differential equation, can no more be used.

In all what follows, the notation (M, ω) will be used te denote as well the *phase space* as the *space of motions* of the dynamical system determined by a smooth Hamiltonian which does not depend on time, according to the context in which it is used.

3.1.2. Definitions

Let $\mathfrak{g}$ be a real, finite-dimensional Lie algebra which acts on a connected symplectic manifold (M, ω) by a Hamiltonian action $\varphi : \mathfrak{g} \to A^1(M)$[11]. Let $J : M \to \mathfrak{g}^*$ be a moment map of the action φ. For each $\boldsymbol{\beta} \in \mathfrak{g}$, we consider the integral

$$\int_M \exp\bigl(-\langle J(x), \boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x)\,, \qquad (*)$$

where λ_ω is the Liouville measure on M.

1. The above integral $(*)$ is said to be *normally convergent* when there exists an open neihbourhood U of $\boldsymbol{\beta}$ in $\mathfrak{g}$ and a function $f : M \to \mathbb{R}^+$, integrable on M with respect to the Liouville measure λ_ω, such that for any $\boldsymbol{\beta}' \in U$, the following inequality

$$\exp\bigl(-\langle J(x), \boldsymbol{\beta}'\rangle\bigr) \leq f(x)$$

is satisfied for all $x \in M$.

2. When $\boldsymbol{\beta} \in \mathfrak{g}$ is such that the integral $(*)$ above is normally convergent, $\boldsymbol{\beta}$ is said to be a *generalized temperature*. The subset of $\mathfrak{g}$ made of generalized temperatures will be denoted by Ω.

3. When the set Ω of generalized temperatures is not empty, the *partition function* associated to the Hamiltonian action φ is the function P defined on Ω by the equality

$$P(\boldsymbol{\beta}) = \int_M \exp\bigl(-\langle J(x), \boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x)\,, \quad x \in M\,, \quad \boldsymbol{\beta} \in \Omega \subset \mathfrak{g}\,.$$

[11]For each $k \in \mathbb{N}$, I denote by $A^k(M)$ the space of fields of k-vectors and by $\Omega^k(M)$ the space of k-exterior differential forms on M, with the convention that $A^0(M) = \Omega^0(M) = C^\infty(M, \mathbb{R})$. For $k = 1$, $A^1(M)$ is therefore the space of smooth vector fields on M. Endowed with the Lie bracket as a composition law, it is an infinite-dimensional Lie algebra.

3.1.3. Proposition

The assumptions and notations are those of Definitions 3.1.2. The set Ω of generalized temperatures does not depend on the choice of the moment map J of the Hamiltonian action φ. When it is not empty, this set is an open convex subset of the Lie algebra $\mathfrak{g}$, the partition function P is of class C^∞ and its differentials of all orders can be calculated by differentiation under the integration sign $\int$.

Proof: When $\boldsymbol{\beta}$ is a generalized temperature, Definitions 3.1.2 imply that there exists a neighbourhood U of $\boldsymbol{\beta}$ whose all elements are generalized temperatures. When it is not empty, the set Ω of generalized temperatures is therefore open. When the moment map J is replaced by another moment map J', the difference $J' - J$ is a constant. The replacement of J by J' has no effect on the eventual normal convergence of the above integral $(*)$, therefore Ω does not depend on the choice of the moment map J.

Let $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ be two distint elements in Ω (assumed to be non-empty), U_0 and U_1 be neighbourhoods, respectively of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$, f_0 and f_1 be the positive functions, defined on M and integrable with respect to the Liouville measure, greater or equal, respectively, than the functions $x \mapsto \exp\bigl(-\langle J(x), \boldsymbol{\beta}'_0\rangle\bigr)$ and $x \mapsto \exp\bigl(-\langle J(x), \boldsymbol{\beta}'_1\rangle\bigr)$ for all $\boldsymbol{\beta}'_0 \in U_0$ and $\boldsymbol{\beta}'_1 \in U_1$. For any $\lambda \in [0,1]$, $U_\lambda = \bigl\{(1-\lambda)\beta'_0 + \lambda\beta'_1 \bigm| \beta'_0 \in U_0\,,\ \beta'_1 \in U_1\bigr\}$ is a neighbourhood of $\boldsymbol{\beta}_\lambda = (1-\lambda)\boldsymbol{\beta}_0 + \lambda\boldsymbol{\beta}_1$. The function $f_\lambda = (1-\lambda)f_0 + \lambda f_1$ is integrable on M. For any $\boldsymbol{\beta}'_\lambda \in U_\lambda$, it is greater or equal to the function $x \mapsto \exp\bigl(-\langle J(x), \boldsymbol{\beta}'_\lambda\rangle\bigr)$. Thereforee $\boldsymbol{\beta}_\lambda \in \Omega$, which proves the convexity of Ω.

For each $x \in M$ fixed, the k-th differential of $\exp\bigl(-\langle J(x), \beta\rangle\bigr)$ with respect to β is

$$D^k\Bigl(\exp\bigl(-\langle J, \beta\rangle\bigr)\Bigr) = (-1)^k J^{\otimes k}(x) \exp\bigl(-\langle J(x), \beta\rangle\bigr)\,,$$

where $J^{\otimes k}(x) = J(x) \otimes \cdots \otimes J(x) \in (\mathfrak{g}^*)^{\otimes k}$. Let us recall that $(\mathfrak{g}^*)^{\otimes k}$ is canonically isomorphic with the space $\mathcal{L}^k(\mathfrak{g}, \mathbb{R})$ of k-multilinear forms on $\mathfrak{g}$. Let us choose any norm on $\mathfrak{g}$. We take on $\mathcal{L}^k(\mathfrak{g}, \mathbb{R})$ the *sup* norm. For any $x \in M$, we have

$$\|J^{\otimes k}(x)\| = \sup_{X_i \in \mathfrak{g}\,,\ \|X_i\| \leq 1\,,\ 1 \leq i \leq k} \bigl|\langle J(x), X_1\rangle \cdots \langle J(x), X_k\rangle\bigr|\,.$$

Let $\boldsymbol{\beta} \in \Omega$ be a generalized temperature. It follows from the definition of a generalized temperature that there exist a real $\varepsilon > 0$ and and a non-negative function f defined on M, integrable with respect to the Liouville measure and greater than the function $x \mapsto \exp\bigl(-\langle J(x), \boldsymbol{\beta}'\rangle\bigr)$ for any $\boldsymbol{\beta}' \in \mathfrak{g}$ satisfying $\|\boldsymbol{\beta}' - \boldsymbol{\beta}\| \leq \varepsilon$. Let $\boldsymbol{\beta}'' \in \mathfrak{g}$ be such that $\|\boldsymbol{\beta}'' - \boldsymbol{\beta}\| \leq \dfrac{\varepsilon}{2}$. For all $X_i \in \mathfrak{g}$ satisfying $\|X_i\| \leq 1$, with

$1 \le i \le k$, and any $x \in M$, we have

$$J^{\otimes k}(x)(X_1, \dots, X_k) = \langle J(x), X_1 \rangle \cdots \langle J(x), X_k \rangle \,.$$

Taking into account the inequality, valid for all $i \in \{1, \dots k\}$,

$$|\langle J(x), X_i \rangle| \le \frac{2k}{\varepsilon} \exp\left(\frac{\varepsilon}{2k} |\langle J(x), X_i \rangle| \right) \,,$$

we can write

$$\begin{aligned} &\Bigl| J^{\otimes k}(x)(X_1, \dots, X_k) \Bigr| \exp\bigl(-\langle J(x), \boldsymbol{\beta}'' \rangle \bigr) \\ &\qquad \le \left(\frac{2k}{\varepsilon} \right)^k \exp\left(- \Bigl\langle J(x), \boldsymbol{\beta}'' + \frac{\varepsilon}{2k}(\eta_1 X_1 + \cdots + \eta_k X_k) \Bigr\rangle \right) \,, \end{aligned}$$

where the terms η_i, $1 \le i \le k$, all equal either to 1 or to -1, are chosen in such a way that $\bigl\langle J(x), \eta_i X_i \bigr\rangle \le 0$. For each $i \in \{1, \dots, k\}$, $|X_i| \le 1$, therefore

$$\left\| \boldsymbol{\beta} - \Bigl(\boldsymbol{\beta}'' + \frac{\varepsilon}{2k}(\eta_1 X_1 + \cdots + \eta_k X_k) \Bigr) \right\| \le \| \boldsymbol{\beta} - \boldsymbol{\beta}'' \| + \frac{\varepsilon k}{2k} \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \,.$$

We see that

$$\Bigl| J^{\otimes k}(x)(X_1, \dots, X_k) \Bigr| \exp\bigl(-\langle J(x), \boldsymbol{\beta}'' \rangle \bigr) \le f(x) \,.$$

By taking the upper bound of the left hand side when the X_i take all possible values among elements in $\mathfrak{g}$ whose norm is smaller than or equal to 1,

$$\Bigl\| J^{\otimes k}(x) \Bigr\| \exp\bigl(-\langle J(x), \boldsymbol{\beta}'' \rangle \bigr) \le f(x) \,.$$

The integral

$$\int_M D^k \Bigl(\exp\bigl(-\langle J(x), \beta \rangle \bigr) \Bigr) \lambda_\omega(\mathrm{d}x)$$

is therefore normally convergent. It follows that the partition function P is of class C^∞, and that its differentials of all orders can be calculated by differentiation under the sign $\int$. ∎

3.1.4. Definition

Let $\boldsymbol{\beta} \in \Omega$ be a generalized temperature. The statistical state on M whose probability density, with respect to the Liouville measure λ_ω, is expressed as

$$\rho_{\boldsymbol{\beta}}(x) = \frac{1}{P(\boldsymbol{\beta})} \exp\bigl(-\langle J(x), \boldsymbol{\beta} \rangle \bigr) \,, \quad x \in M \,,$$

is called the *Gibbs state* associated to (or indexed by) $\boldsymbol{\beta}$.

3.1.5. Proposition

For any generalized temperature $\boldsymbol{\beta} \in \Omega$, the integral below

$$\mathcal{E}_{\rho_{\boldsymbol{\beta}}}(J) = \frac{1}{P(\boldsymbol{\beta})} \int_M J(x) \exp\bigl(- \langle J(x), \boldsymbol{\beta} \rangle \bigr) \lambda_\omega(\mathrm{d}x)$$

is convergent. This integral defines the mean value $\mathcal{E}_{\rho_{\boldsymbol{\beta}}}(J)$ of the moment map J in the Gibbs state indexed by $\boldsymbol{\beta}$. Moreover, for any other continuous statistical state with a probability density ρ_1 with respect to the Liouville measure λ_ω, such that $\mathcal{E}_{\rho_1}(J)$ exists and is equal to $\mathcal{E}_{\rho_{\boldsymbol{\beta}}}(J)$, the entropy functional s satisfies the inequality $s(\rho_1) \leq s(\rho_\beta)$, and the equality $s(\rho_1) = s(\rho_\beta)$ occurs if and only if $\rho_1 = \rho_\beta$.

Proof: The normal convergence (which implies the usual convergence) of the integral which defines $\mathcal{E}_{\rho_{\boldsymbol{\beta}}}(J)$ follows from Proposition 3.1.3. Let ρ_1 be the probability density, with respect to λ_ω, of another continuous statistical state such that $\mathcal{E}_{\rho_1}(J)$ exists and is equal to $\mathcal{E}_{\rho_{\boldsymbol{\beta}}}(J)$. The function, defined on $\mathbb{R}^+$,

$$z \mapsto h(z) = \begin{cases} z \log \left(\dfrac{1}{z} \right) & \text{if } z > 0 \\ 0 & \text{if } z = 0 \end{cases}$$

being convex, the straight line tangent to its graph at one of its point $\bigl(z_0, h(z_0)\bigr)$ is always above this graph. Therefore, for all $z > 0$ and $z_0 > 0$, the following inequality holds:

$$h(z) \leq h(z_0) - (1 + \log z_0)(z - z_0) = z_0 - z(1 + \log z_0)\,.$$

With $z = \rho_1(x)$ and $z_0 = \rho_{\boldsymbol{\beta}(}x)$, for any $x \in M$, this inequality becomes

$$h\bigl(\rho_1(x)\bigr) = \rho_1(x) \log \left(\frac{1}{\rho_1(x)} \right) \leq \rho_{\boldsymbol{\beta}}(x) - \bigl(1 + \log \rho_{\boldsymbol{\beta}}(x)\bigr) \rho_1(x)\,.$$

By integrating on M both sides of the above inequality, we get, since $\rho_{\boldsymbol{\beta}}$ is the probability density of the Gibbs state indexed by $\boldsymbol{\beta}$,

$$s(\rho_1) \leq 1 - 1 - \int_M \rho_1(x) \log \rho_{\boldsymbol{\beta}}(x) \lambda_\omega(\mathrm{d}x) = s(\rho_\beta)\,.$$

We have proven the inequality $s(\rho_1) \leq s(\rho_\beta)$. If $\rho_1 = \rho_\beta$, of course $s(\rho_1) = s(\rho_\beta)$. Conversely, let us now assume that $s(\rho_1) = s(\rho_\beta)$. The functions φ_1 and φ, defined on M, whose expressions are

$$\varphi_1(x) = \rho_1(x) \log \left(\frac{1}{\rho_1(x)} \right)\,,\ \varphi(x) = \rho_{\boldsymbol{\beta}}(x) - \bigl(1 + \log \rho_{\boldsymbol{\beta}}(x)\bigr) \rho_1(x)\,,\ x \in M\,,$$

are continuous, except, maybe, the function φ at points x where $\rho_{\boldsymbol{\beta}}(x) = 0$ and $\rho_1(x) \neq 0$. For the Liouville measure λ_ω, the subset of M made of these points is of measure 0, since φ is integrable. The functions φ and φ_1 satisfy the inequality $\varphi_1 \leq \varphi$, are integrable on M and their integrals are equal. Their difference $\varphi - \varphi_1$, everywhere ≥ 0 on M and with an integral equal to 0, is therefore everywhere equal to 0. Therefore, for any $x \in M$,

$$\rho_1(x) \log\left(\frac{1}{\rho_1(x)}\right) = \rho_{\boldsymbol{\beta}}(x) - \big(1 + \log \rho_{\boldsymbol{\beta}}(x)\big)\rho_1(x)\,. \tag{$*$}$$

For any $x \in M$ such that $\rho_1(x) \neq 0$, we can divide both sides of the above equality by $\rho_1(x)$. We get

$$\frac{\rho_\beta(x)}{\rho_1(x)} - \log\left(\frac{\rho_{\boldsymbol{\beta}}(x)}{\rho_1(x)}\right) = 1\,.$$

The function $z \mapsto z - \log z$ reaches its minimum at only one point $z > 0$, the point $z = 1$, and its minimum is equal to 1. So for all $x \in M$ such that $\rho_1(x) > 0$, $\rho_1(x) = \rho_{\boldsymbol{\beta}}(x)$. At points $x \in M$ such that $\rho_1(x) = 0$, equality $(*)$ proves that $\rho_{\boldsymbol{\beta}}(x) = 0$. Therefore $\rho_1 = \rho_{\boldsymbol{\beta}}$ everywhere on M. ■

3.1.6. Proposition

We now assume that $\mathfrak{g}$ is the Lie algebra of a Lie group G which acts on the symplectic manifold (M,ω) by a Hamiltonian action Φ, and that φ is the Lie algebra action associated to Φ. The Gibbs state indexed by any generalized temperature $\boldsymbol{\beta} \in \Omega$ is invariant by the restriction of the action Φ to the one-parameter subgroup $\big\{\exp(\tau\beta) \bigm| \tau \in \mathbb{R}\big\}$ of G.

Proof: The orbits of the action of this subgroup on M are the integral curves of the Hamiltonian vector field which admits the function $x \mapsto \big\langle J(x), \boldsymbol{\beta}\big\rangle$ as Hamiltonian. This function is therefore constant on each orbit of this one-parameter subgroup. The expression of the probability density $\rho_{\boldsymbol{\beta}}$ of the Gibbs state indexed by $\boldsymbol{\beta}$ proves that this probability density too is constant on each orbit of the action of $\big\{\exp(\tau\beta) \bigm| \tau \in \mathbb{R}\big\}$. ■

3.2. Thermodynamic functions

In this section, the map $\Phi : G \times M \to M$ is a Hamiltonian action of a Lie group G on a connected symplectic manifold (M,ω) and $J : M \to \mathfrak{g}^*$ is a moment map of this action. It is assumed that the open subset $\Omega \subset \mathfrak{g}$ of generalized temperatures is not empty. In Definitions 3.1.2, we have defined the partition function P, whose expression is

$$P(\boldsymbol{\beta}) = \int_M \exp\bigl(-\langle J(x), \beta\rangle\bigr)\lambda_\omega(\mathrm{d}x)\,, \quad \boldsymbol{\beta} \in \Omega\,.$$

On the set Ω of generalized temperatures, we define below other *thermodynamic functions* whose expressions can be derived from that of P.

3.2.1. Definitions

Assumptions and notations here are those of Subsection 3.2.

1. The *mean value of the moment map* J is the function, denoted by E_J, defined on Ω and taking its values in the dual vector space $\mathfrak{g}^*$ of the Lie algebra $\mathfrak{g}$, whose expression is

$$E_J(\boldsymbol{\beta}) = \mathcal{E}_{\rho_{\boldsymbol{\beta}}}(J) = \frac{1}{P(\boldsymbol{\beta})}\int_M J(x)\exp\bigl(-\langle J(x), \boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x)\,, \quad \boldsymbol{\beta} \in \Omega\,.$$

2. The *entropy function* is the function, denoted by S, defined on Ω and taking its values in $\mathbb{R}\cup\{-\infty\}$, which associates, to each generalized temperature $\boldsymbol{\beta} \in \Omega$, the entropy of the Gibbs state indexed by $\boldsymbol{\beta}$:

$$S(\boldsymbol{\beta}) = s(\rho_{\boldsymbol{\beta}}) = \int_M \rho_{\boldsymbol{\beta}}(x)\log\left(\frac{1}{\rho_{\boldsymbol{\beta}}(x)}\right)\lambda_\omega(\mathrm{d}x)\,, \quad \boldsymbol{\beta} \in \Omega\,.$$

3.2.2. Proposition

For each generalized temperature $\boldsymbol{\beta} \in \Omega$, the values at $\boldsymbol{\beta}$ of the thermodynamic functions mean value of J and entropy (Definitions 3.2.2) are given by the formulae

$$\begin{aligned} E_J(\boldsymbol{\beta}) &= -\frac{1}{P(\boldsymbol{\beta})}DP(\boldsymbol{\beta}) = -D(\log P)(\boldsymbol{\beta})\,,\\ S(\boldsymbol{\beta}) &= \log P(\boldsymbol{\beta}) + \bigl\langle E_J(\boldsymbol{\beta}), \boldsymbol{\beta}\bigr\rangle = \log P(\boldsymbol{\beta}) - \bigl\langle D(\log P)(\boldsymbol{\beta}), \boldsymbol{\beta}\bigr\rangle\,. \end{aligned}$$

Proof: Proposition 3.1.3 states that the partition function P is of class C^∞ and that its differentials of all orders can be obtained by differentiation under the sign $\int$. Therefore

$$DP(\boldsymbol{\beta}) = -\int_M J\exp\bigl(-\langle J(x), \boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x) = -P(\boldsymbol{\beta})E_J(\boldsymbol{\beta})\,,$$

which proves the indicated expresions of $E_J(\beta)$. Since for each $x \in M$, we have $\rho_{\boldsymbol{\beta}}(x) = \dfrac{\exp\bigl(-\langle J(x), \boldsymbol{\beta}\rangle\bigr)}{P(\boldsymbol{\beta})}$,

$$\rho_{\boldsymbol{\beta}}(x) \log \frac{1}{\rho_{\boldsymbol{\beta}}(x)} = \frac{1}{P(\boldsymbol{\beta})} \exp\bigl(-\langle J(x), \boldsymbol{\beta}\rangle\bigr)\bigl(\langle J(x), \boldsymbol{\beta}\rangle + \log P(\boldsymbol{\beta})\bigr) .$$

By integration over M of both members of this equality with respect to λ_ω, we obtain the indicated expression of $S(\boldsymbol{\beta})$. ∎

3.2.3. Proposition

The thermodynamic functions E_J (mean value of J) and S (entropy) are of class C^∞ on Ω. The first differential of E_J is the function, defined on Ω and taking its values in the space of linear applications of $\mathfrak{g}$ in its dual vector space $\mathfrak{g}^*$, whose expression is

$$\bigl\langle DE_J(\boldsymbol{\beta})(X), Y\bigr\rangle = -D^2(\log P)(\boldsymbol{\beta})(X, Y)$$
$$= -\frac{1}{P(\boldsymbol{\beta})} \int_M \bigl\langle J(x) - E_J(\boldsymbol{\beta}), X\bigr\rangle \bigl\langle J(x) - E_J(\boldsymbol{\beta}), Y\bigr\rangle \exp\bigl(-\langle J(x), \boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x) ,$$

with X and $Y \in \mathfrak{g}$. For each $\boldsymbol{\beta} \in \Omega$, $DE_J(\boldsymbol{\beta})$ can be considered as a bilinear, symmetric form on $\mathfrak{g}$.

The differential of the entropy function S at each $\boldsymbol{\beta} \in \Omega$ is an element of $\mathfrak{g}^*$ whose expression is

$$\bigl\langle DS(\boldsymbol{\beta}), X\bigr\rangle = \bigl\langle DE_J(\boldsymbol{\beta})(X), \boldsymbol{\beta}\bigr\rangle , \quad X \in \mathfrak{g} .$$

Proof: According to Proposition 3.2.2, for each $\boldsymbol{\beta} \in \Omega$, $E_J(\boldsymbol{\beta}) = -D(\log P)(\boldsymbol{\beta})$. Therefore, for all X and $Y \in \mathfrak{g}$,

$$\bigl\langle DE_J(\boldsymbol{\beta})(X), Y\bigr\rangle = -D^2(\log P)(\boldsymbol{\beta})(X, Y) ,$$

which shows that $DE_J(\boldsymbol{\beta})$ can be considered as a bilinear, symmetric form on $\mathfrak{g}$. Since $DP(\boldsymbol{\beta})$ can be obtained by differentiation under the sign $\int$,

$$D(\log P)(\boldsymbol{\beta}) = \frac{1}{P(\boldsymbol{\beta})} DP(\boldsymbol{\beta}) = -\frac{1}{P(\boldsymbol{\beta})} \int_M J(x) \exp\bigl(-\langle J(x), \boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x) .$$

By a second differentiation under the sign $\int$, we therefore obtain, for all X and $Y \in \mathfrak{g}$,

$$D^2(\log P)(\boldsymbol{\beta})(X, Y) = \frac{1}{P(\boldsymbol{\beta})} \int_M \langle J(x), X\rangle \langle J(x), Y\rangle \exp\bigl(-\langle J(x), \boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x)$$
$$+ \frac{1}{\bigl(P(\boldsymbol{\beta})\bigr)^2} DP(\boldsymbol{\beta})(Y) \int_M \langle J(x), X\rangle \exp\bigl(-\langle J(x), \boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x) .$$

Let us replace $DP(\boldsymbol{\beta})(Y)$ and $\int_M \langle J(x), X\rangle \exp\bigl(-\langle J(x),\boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x)$, in the right hand side of this equality, by their expressions

$$DP(\boldsymbol{\beta})(Y) = -P(\boldsymbol{\beta})\bigl\langle E_J(\boldsymbol{\beta}), Y\bigr\rangle\,,$$
$$\int_M \langle J(x), X\rangle \exp\bigl(-\langle J(x),\boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x) = P(\boldsymbol{\beta})\bigl\langle E_J(\boldsymbol{\beta}, X\bigr\rangle\,.$$

We obtain

$$\begin{aligned} D^2(\log P)(\boldsymbol{\beta})(X,Y) &= \frac{1}{P(\boldsymbol{\beta})}\int_M \langle J(x), X\rangle\langle J(x), Y\rangle \exp\bigl(-\langle J(x),\boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x)\\ &\quad -\bigl\langle E_J(\boldsymbol{\beta}), Y\bigr\rangle\bigl\langle E_J(\boldsymbol{\beta}, X\bigr\rangle\\ &= \frac{1}{P(\boldsymbol{\beta})}\int_M \bigl\langle J(x)-E_J(\boldsymbol{\beta}), X\bigr\rangle\bigl\langle J(x)-E_J(\boldsymbol{\beta}), Y\bigr\rangle\\ &\qquad\qquad \exp\bigl(-\langle J(x),\boldsymbol{\beta}\rangle\bigr)\lambda_\omega(\mathrm{d}x)\,, \end{aligned}$$

The expression of $\bigl\langle DE_J(\boldsymbol{\beta})(X), Y\bigr\rangle$ given in the statement follows.

By differentiation of the expression of $S(\boldsymbol{\beta})$ given in Proposition 3.2.2, and by using the equality $DP(\boldsymbol{\beta}) = -P(\boldsymbol{\beta})E_J(\boldsymbol{\beta})$, we obtain, for any $X\in\mathfrak{g}$,

$$\begin{aligned} \bigl\langle DS(\boldsymbol{\beta}), X\bigr\rangle &= \frac{1}{P(\boldsymbol{\beta})}DP(\boldsymbol{\beta})(X) + \bigl\langle E_J(\boldsymbol{\beta}), X\bigr\rangle + \bigl\langle DE_J(\boldsymbol{\beta})(X), \boldsymbol{\beta}\bigr\rangle\\ &= \bigl\langle DE_J(\boldsymbol{\beta})(X), \boldsymbol{\beta}\bigr\rangle\,. \end{aligned}$$

■

3.2.4. Theorem

For all $\boldsymbol{\beta}\in\Omega$, X and $Y\in\mathfrak{g}$, let

$$\Gamma(\boldsymbol{\beta})(X,Y) = -\bigl\langle DE_J(\boldsymbol{\beta})(X), Y\bigr\rangle = D^2(\log P)(\boldsymbol{\beta})(X,Y)\,.$$

The map Γ so defined is a C^∞ bilinear, symmetric differential form defined on Ω such that, for each $\boldsymbol{\beta}\in\Omega$ and $X\in\mathfrak{g}$,

$$\Gamma(\boldsymbol{\beta})(X,X)\geq 0\,.$$

Moreover, if $X\in\mathfrak{g}$ is such that $x\mapsto\bigl\langle J(x), X\bigr\rangle$ is not a constant function,

$$\Gamma(\boldsymbol{\beta})(X,X) > 0\,.$$

When, in addition, the Hamiltonian action $\Phi: G\times M\to M$ is effective (it means that for any $X\in\mathfrak{g}$, $X\neq 0$, the function $x\mapsto\bigl\langle J(x), X\bigr\rangle$ is not a constant on

M), Γ is a Riemannian metric on Ω. Moreover, the map $E_J : \Omega \to \mathfrak{g}^*$ is injective, its image is an open subset Ω^* of $\mathfrak{g}^*$, and considered as valued in Ω^*, E_J is a diffeomorphism of the set Ω of generalized temperatures onto the open subset Ω^* of $\mathfrak{g}^*$.

Proof: The firt assertions follow from the the expression of $\bigl\langle DE_J(\boldsymbol{\beta})(X), Y\bigr\rangle$ given in Proposition 3.2.3. When $X \in \mathfrak{g}$ is such that the function $x \mapsto \bigl\langle J(x), X\bigr\rangle$ is not a constant on M, the function

$$x \mapsto \bigl\langle J(x) - E_J(\boldsymbol{\beta}), X\bigr\rangle^2 \exp\bigl(-\langle J, \boldsymbol{\beta}\rangle\bigr)$$

is continuous, with values ≥ 0 and not everywhere equal to 0 on M. Its integral with respect to the Liouville measure is therefore strictly positive, which proves that $\Gamma(\boldsymbol{\beta})(X, X) > 0$.

When, in addition, the action Φ effective, for any $\boldsymbol{\beta} \in \Omega$ and any non-zero $X \in \mathfrak{g}$, $\Gamma(\boldsymbol{\beta})(X, X) > 0$. The map Γ is therefore an Riemannian metric on Ω. For all $\boldsymbol{\beta} \in \Omega$ and $Y \in \mathfrak{g}\backslash\{0\}$, we have $\bigl\langle DE_J(\boldsymbol{\beta})(Y), Y\bigr\rangle < 0$, which implies that $DE_J(\boldsymbol{\beta})$ is invertible. The map $E_J : \Omega \to \mathfrak{g}^*$ is therefore open. This map cannot take the same value at two distinct points $\boldsymbol{\beta}_1$ an $\boldsymbol{\beta}_2 \in \Omega$, since this would imply

$$\bigl\langle E_J(\boldsymbol{\beta}_1), \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1\bigr\rangle = \bigl\langle E_J(\boldsymbol{\beta}_2), \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1\bigr\rangle\,.$$

The real-valued function

$$\lambda \mapsto \Bigl\langle E_J\bigl((1-\lambda)\boldsymbol{\beta}_1 + \lambda\boldsymbol{\beta}_2\bigr), \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1\Bigr\rangle\,, \quad \lambda \in [0, 1]\,,$$

would be well defined on $[0, 1]$ since Ω is convex, smooth in $]0, 1[$, and would take the same value for $\lambda = 0$ and $\lambda = 1$. Its derivative with respect to λ, whose value is $\Bigl\langle DE_J\bigl(\lambda\boldsymbol{\beta}_1 + (1-\lambda)\boldsymbol{\beta}_2\bigr)(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1), (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1)\Bigr\rangle$ would vanish fore some $\lambda \in]0, 1[$, which would contradict the effectiveness of Φ. Being open and injective, the map $E_J : \Omega \to \mathfrak{g}^*$ is a diffeomorphism of Ω onto its image Ω^*, which is an open subset of $\mathfrak{g}^*$. ■

3.2.5. Remarks

Theorem 3.2.4 leads to the following observations.

1. In the language of Probability theory, $-\bigl\langle DE_J(\boldsymbol{\beta})(X), X\bigr\rangle$ is the *variance*, in other words the square of the *standard deviation* of the random variable $\langle J, X\rangle$, for the probability law $\rho_{\boldsymbol{\beta}}\lambda_\omega$ on M.

2. For each generalized temperature $\boldsymbol{\beta} \in \Omega$, the Gibbs state indexed by $\boldsymbol{\beta}$ is the probability law on M, absolutely continuous with respect to the Liouville measure λ_ω, of probability density

$$\rho_{\boldsymbol{\beta}} = \frac{1}{P(\boldsymbol{\beta})} \exp\bigl(-\langle J, \boldsymbol{\beta}\rangle\bigr)\,.$$

The open subset Ω of $\mathfrak{g}$, in which live the generalized temperatures $\boldsymbol{\beta}$ which index a familly of probability laws defined on M, is called by statisticians a *statistical manifold*. The *Fisher-Rao metric*, so named in honour of the British statistician and genetician Ronald Aylmer Fisher (1890–1962) and the Indian statistician Calyampudi Radhakrishna Rao (born in 1920, emeritus Professor at the Indian Statistics Institute and at the Pennsylvania State University) is a Riemannian metric, defined on some statistical manifolds, which is used to evaluate the distance between probability laws. Frédéric Barbaresco [5] observed that the Riemannian metric Γ defined by Jean-Marie Souriau on Ω is nothing else than the Fisher-Rao metric when, as indicated above, Ω is considered as a statistical manifold. He observed too that this metric already appeared in the works of the French mathematician René Maurice Fréchet (1878–1973) [11], two years before it was rediscovered by Fisher and Rao.

3. Under the assumptions of Theorem 3.2.4, the equality

$$S(\boldsymbol{\beta}) = \bigl\langle D(-\log P)(\boldsymbol{\beta}), \boldsymbol{\beta}\bigr\rangle - (-\log P)(\boldsymbol{\beta})$$

proves that each of the two functions $-\log P : \Omega \to \mathbb{R}$ and $S \circ E_J^{-1} : \Omega^* \to \mathbb{R}$ is the *Legendre transform* of the other, just as a hyper-regular Lagrangian $L : TM \to \mathbb{R}$ and the associated Hamiltonian $H : T^*M \to \mathbb{R}$, defined, respectively, on the tangent bundle TM and on the cotangent bundle T^*M to some smooth manifold M. Here the Legendre map is $E_J : \Omega \to \Omega^*$. This map and its inverse $(E_J)^{-1} : \Omega^* \to \Omega$ are expressed by formulae similar to those which express the Legendre map $TM \to T^*M$ and its inverse $T^*M \to TM$ in calculus of variations,

$$E_J = D(-\log P)\,, \quad (E_J)^{-1} = D(S \circ {E_J}^{-1})\,.$$

The moment map J of the Hamiltonian action Φ is not unique: it is well known that for any constant $\mu \in \mathfrak{g}$, $J + \mu$ is too a moment map of Φ. Proposition 3.2.6 below indicates the effet of such a change on the thermodynamic functions P, E_J and S.

3.2.6. Proposition

Let $\mu \in \mathfrak{g}^*$ be a constant. When the moment map J of the Hamiltonian action Φ is replaced by $J_1 = J + \mu$, the set Ω of generalized temperatures does not change. The thermodynamic functions P, E_J and S are replaced, respectively, by P_1, E_{J_1} and S_1, whose expressions are

$$P_1(\boldsymbol{\beta}) = \exp\bigl(-\langle\mu, \boldsymbol{\beta}\rangle\bigr)P(\boldsymbol{\beta})\,, \quad E_{J_1}(\boldsymbol{\beta}) = E_J(\boldsymbol{\beta}) + \mu\,, \quad S_1(\boldsymbol{\beta}) = S(\boldsymbol{\beta})\,.$$

For each $\boldsymbol{\beta} \in \Omega$, the associated Gibbs state, its probability density $\rho_{\boldsymbol{\beta}}$ with respect to the Liouville measure λ_ω and the bilinear, symmetric form Γ (Theorem 3.2.4) are not changed.

Proof: The stated results follow from the equality

$$\exp\bigl(-\langle J + \mu, \boldsymbol{\beta}\rangle\bigr) = \exp\bigl(-\langle\mu, \boldsymbol{\beta}\rangle\bigr)\exp\bigl(-\langle J, \boldsymbol{\beta}\rangle\bigr)\,.$$

■

3.3. Generalized temperatures and adjoint action

As in the previous section, $\Phi : G \times M \to M$ is a Hamiltonian action of a connected Lie group G on a connected symplectic manifold (M, ω) and $J : M \to \mathfrak{g}^*$ is a moment map of this action. The set of generalized temperatures is assumed to be a non-empty subset Ω of the Lie algebra $\mathfrak{g}$. As seen in Proposition 3.2.6, Ω does not depend on the choice of the moment map J. We moreover assume that Φ is effective, which implies (Theorem 3.2.4) that E_J is a diffeomorphism of Ω onto an open subset Ω^* of $\mathfrak{g}^*$, and that the bilinear, symmetric form Γ is a Riemannian metric on Ω. By considering the adjoint action of G on Ω, we prove below that Ω is a union of adjoint orbits (Proposition 3.3.1) and that the Riemannian metric induced by Γ on each of these orbits can be expressed in terms of a symplectic cocycle of the Lie algebra $\mathfrak{g}$ (Theorem 3.3.4).

The next proposition proves that Ω is a union of adjoint orbits and indicates the variations of the thermodynamic functions P, E_J and S on each adjoint orbit contained in Ω.

3.3.1. Proposition

The set Ω of generalized temperatures is a union of orbits of the adjoint action of the Lie group G on its Lie algebra $\mathfrak{g}$. Let $\theta : G \to \mathfrak{g}^*$ be the symplectic cocycle of G (see, for example, [18]) such that, for each $g \in G$

$$J \circ \Phi_g = \mathrm{Ad}^*_{g^{-1}} \circ J + \theta(g)\,.$$

For any $\boldsymbol{\beta} \in \Omega$ and any $g \in G$, we have

$$\begin{aligned} P(\mathrm{Ad}_g\boldsymbol{\beta}) &= \exp\Big(\big\langle\theta(g^{-1}),\boldsymbol{\beta}\big\rangle\Big)P(\boldsymbol{\beta}) = \exp\Big(-\big\langle \mathrm{Ad}_g^*\theta(g),\boldsymbol{\beta}\big\rangle\Big)P(\boldsymbol{\beta})\,,\\ E_J(\mathrm{Ad}_g\boldsymbol{\beta}) &= \mathrm{Ad}_{g^{-1}}^* E_J(\boldsymbol{\beta}) + \theta(g)\,,\\ S(\mathrm{Ad}_g\boldsymbol{\beta}) &= S(\boldsymbol{\beta})\,. \end{aligned}$$

Proof: Let us assume that the integral which defines $P(\mathrm{Ad}_g\boldsymbol{\beta})$ is convergent. We can write

$$\begin{aligned} P(\mathrm{Ad}_g\boldsymbol{\beta}) &= \int_M \exp\big(-\langle J(x),\mathrm{Ad}_g\boldsymbol{\beta}\rangle\big)\lambda_\omega(\mathrm{d}x)\\ &= \int_M \exp\big(-\langle \mathrm{Ad}_g^* J(x),\boldsymbol{\beta}\rangle\big)\lambda_\omega(\mathrm{d}x)\\ &= \int_M \exp\Big(-\big\langle J\circ\Phi_{g^{-1}}(x)-\theta(g^{-1}),\boldsymbol{\beta}\big\rangle\Big)\lambda_\omega(\mathrm{d}x)\\ &= \exp\Big(\big\langle\theta(g^{-1}),\boldsymbol{\beta}\big\rangle\Big)\int_M \exp\Big(-\big\langle J\circ\Phi_{g^{-1}}(x),\boldsymbol{\beta}\big\rangle\Big)\lambda_\omega(\mathrm{d}x)\,. \end{aligned}$$

The change of integration variable $y = \Phi_{g^{-1}}(x)$ in the last integral leads to

$$\int_M \exp\Big(-\big\langle J\circ\Phi_{g^{-1}}(x),\boldsymbol{\beta}\big\rangle\Big)\lambda_\omega(\mathrm{d}x) = \int_M \exp\Big(-\big\langle J(y),\boldsymbol{\beta}\big\rangle\Big)\Phi_g^*\lambda_\omega(\mathrm{d}y) = P(\boldsymbol{\beta})\,,$$

since $\Phi_g^*\lambda_\omega = \lambda_\omega$, the Liouville measure being invariant by symplectomophisms. Moreover, $\theta(g^{-1}) = -\mathrm{Ad}_g^*\theta(g)$ (see for example [18]), so we can write

$$P(\mathrm{Ad}_g\boldsymbol{\beta}) = \exp\Big(-\big\langle \mathrm{Ad}_g^*\theta(g),\boldsymbol{\beta}\big\rangle\Big)P(\boldsymbol{\beta})\,.$$

By reversing the above calculation step by step, we prove that the normal convergence of the integral which defines $P(\boldsymbol{\beta})$ implies the normal convergence of the integral which defines $P(\mathrm{Ad}_g\boldsymbol{\beta})$. We therefore have proven that Ω is a union of adjoint orbits of G, as well as the expression of $P(\mathrm{Ad}_g\boldsymbol{\beta})$ in terms of $P(\boldsymbol{\beta})$ and θ given in the statement.

Since $E_J(\boldsymbol{\beta}) = -D(\log P)(\boldsymbol{\beta})$, $E_J(\mathrm{Ad}_g\boldsymbol{\beta}) = -D(\log P)(\mathrm{Ad}_g\boldsymbol{\beta})$. To calculate the right hand side of this equality, we observe that for each $\boldsymbol{\delta} \in \mathfrak{g}$ and each real s,

$$\begin{aligned} D(\log P)(\mathrm{Ad}_g\boldsymbol{\beta})(\boldsymbol{\delta}) &= \frac{\mathrm{d}}{\mathrm{d}s}\big(\log P(\mathrm{Ad}_g\boldsymbol{\beta}+s\boldsymbol{\delta})\big)\,\big|_{s=0}\\ &= \frac{\mathrm{d}}{\mathrm{d}s}\Big(\log P\big(\mathrm{Ad}_g(\boldsymbol{\beta}+s\mathrm{Ad}_{g^{-1}}\boldsymbol{\delta})\big)\Big)\,\big|_{s=0} \end{aligned}$$

Using the expression of $P(\mathrm{Ad}_g\boldsymbol{\beta})$ obtained above, we have

$$\log P\bigl(\mathrm{Ad}_g(\boldsymbol{\beta}+s\mathrm{Ad}_{g^{-1}}\boldsymbol{\delta})\bigr) = -\langle \mathrm{Ad}_g^*\theta(g), \boldsymbol{\beta}+s\mathrm{Ad}_{g^{-1}}\boldsymbol{\delta}\rangle + \log P(\boldsymbol{\beta}+s\mathrm{Ad}_{g^{-1}}\boldsymbol{\delta})\,.$$

Taking the derivative with respect to s, then setting $s=0$, we get

$$\begin{aligned} D\log P(\mathrm{Ad}_g\boldsymbol{\beta})(\boldsymbol{\delta}) &= -\bigl\langle \mathrm{Ad}_g^*\theta(g), \mathrm{Ad}_{g^{-1}}\boldsymbol{\delta}\bigr\rangle + D\log P(\boldsymbol{\beta})(\mathrm{Ad}_{g^{-1}}\boldsymbol{\delta})\\ &= -\bigl\langle \theta(g), \boldsymbol{\delta}\bigr\rangle + D\log(P)(\boldsymbol{\beta})(\mathrm{Ad}_{g^{-1}}\boldsymbol{\delta})\\ &= -\bigl\langle \theta(g) + \mathrm{Ad}_{g^{-1}}^* E_J(\boldsymbol{\beta}), \boldsymbol{\delta}\bigr\rangle\,, \end{aligned}$$

where we have used the already obtained equality $D\log P(\boldsymbol{\beta}) = -E_J(\boldsymbol{\beta})$. Therefore,

$$E_J(\mathrm{Ad}_g\boldsymbol{\beta}) = \mathrm{Ad}_{g^{-1}}^* E_J(\boldsymbol{\beta}) + \theta(g)\,.$$

Finally,

$$\begin{aligned} S(\mathrm{Ad}_g\boldsymbol{\beta}) &= \log P(\mathrm{Ad}_g\boldsymbol{\beta}) - \bigl\langle D\log P(\mathrm{Ad}_g\boldsymbol{\beta}), \mathrm{Ad}_g\boldsymbol{\beta}\bigr\rangle\\ &= -\bigl\langle \mathrm{Ad}_g^*\theta(g), \boldsymbol{\beta}\bigr\rangle + \log P(\boldsymbol{\beta}) + \bigl\langle \mathrm{Ad}_{g^{-1}}^* E_J(\boldsymbol{\beta}) + \theta(g), \mathrm{Ad}_g\boldsymbol{\beta}\bigr\rangle\\ &= \log P(\boldsymbol{\beta}) + \bigl\langle E_J(\boldsymbol{\beta}), \boldsymbol{\beta}\bigr\rangle\\ &= S(\boldsymbol{\beta})\,. \end{aligned}$$

■

3.3.2. Remark

The equality

$$E_J(\mathrm{Ad}_g\boldsymbol{\beta}) = \mathrm{Ad}_{g^{-1}}^* E_J(\boldsymbol{\beta}) + \theta(g)$$

states that the map $E_J : \Omega \to \Omega^*$ is equivariant with respect to the adjoint action Φ of G on $\mathfrak{g}$, restricted to the open subset Ω of $\mathfrak{g}$, and its affine action a_θ on $\mathfrak{g}^*$:

$$a_\theta(g,\xi) = \mathrm{Ad}_{g^{-1}}^*\xi + \theta(g)\,,\quad g\in G\,,\quad \xi\in\mathfrak{g}^*\,,$$

restricted to the open subset Ω^* of $\mathfrak{g}^*$. This result is not surprising, since it is well known (see, for example, [18]) that the moment map J itself is equivariant with respect to the action Φ of G on M and its affine action a_θ on $\mathfrak{g}^*$: it states that the equivariance of J implies the equivariance of its mean value.

3.3.3. Proposition

Let $\Theta = T_e\theta : \mathfrak{g} \to \mathfrak{g}^*$ be the 1*-cocycle of the Lie algebra* $\mathfrak{g}$ *associed to the symplectic* 1*-cocycle* θ *of the Lie group* G *(see, for example, [18]). For any* $\boldsymbol{\beta} \in \Omega$ *and any* $X \in \mathfrak{g}$,

$$\begin{aligned}\bigl\langle E_J(\boldsymbol{\beta}), [X, \boldsymbol{\beta}]\bigr\rangle &= \bigl\langle \Theta(X), \boldsymbol{\beta}\bigr\rangle\,,\\ DE_J(\boldsymbol{\beta})\bigl([X, \boldsymbol{\beta}]\bigr) &= -\mathrm{ad}^*_X E_J(\boldsymbol{\beta}) + \Theta(X)\,.\end{aligned}$$

Proof: Let us set $g = \exp(\tau X)$ in the expression of $P(\mathrm{Ad}_g\boldsymbol{\beta})$ given in Proposition 3.3.1, then take the derivative with respect to τ and set $\tau = 0$. Using the well known equalities $\theta(e) = 0$ and $T_e\theta = \Theta$, we obtain

$$\begin{aligned}DP(\boldsymbol{\beta})\bigl([X, b]\bigr) &= \frac{\mathrm{d}}{\mathrm{d}\tau}\Bigl(\exp\bigl(-\bigl\langle \mathrm{Ad}^*_{\exp(\tau X)}\theta\bigl(\exp(\tau X)\bigr), \boldsymbol{\beta}\bigr\rangle\bigr)P(\boldsymbol{\beta})\Bigr)\Bigr|_{\tau=0}\\ &= -\bigl\langle \Theta(X), \boldsymbol{\beta}\bigr\rangle P(\boldsymbol{\beta})\,,\end{aligned}$$

which proves the first assertion, since $DP(\boldsymbol{\beta}) = -P(\boldsymbol{\beta})E_J(\boldsymbol{\beta})$.

Similarly, let us set $g = \exp(\tau X)$ in the expression of $E_J(\mathrm{Ad}_g\boldsymbol{\beta})$ given in Proposition 3.3.1, then take the derivative with respect to τ and set $\tau = 0$. We obtain

$$\begin{aligned}DE_J(\boldsymbol{\beta})\bigl([X, \boldsymbol{\beta}]\bigr) &= \frac{\mathrm{d}}{\mathrm{d}\tau}\Bigl(\mathrm{Ad}^*_{\exp(-\tau X)}E_J(\boldsymbol{\beta}) + \theta\bigl(\exp(\tau X)\bigr)\Bigr)\Bigr|_{\tau=0}\\ &= -\mathrm{ad}^*_X E_J(\boldsymbol{\beta}) + \Theta(X)\,.\end{aligned}$$

■

3.3.4. Theorem

Let us set, for each generalized temperature $\boldsymbol{\beta} \in \Omega$,

$$J_{\boldsymbol{\beta}} = J - E_J(\boldsymbol{\beta})\,,$$

and, for each $g \in G$,

$$\theta_{\boldsymbol{\beta}}(g) = \theta(g) - E_J(\boldsymbol{\beta}) + \mathrm{Ad}^*_{g^{-1}}E_J(\boldsymbol{\beta})\,.$$

The map $J_{\boldsymbol{\beta}}$ *is the unique moment map of the Hamiltonian action* Φ *whose mean value, for the generalized temperature* $\boldsymbol{\beta}$*, is equal to* 0*. The map* $\theta_{\boldsymbol{\beta}}$ *is the symplectic cocycle of the Lie group* G*, cohomologous to* θ*, associated to the moment map* $J_{\boldsymbol{\beta}}$*. It depends on* $\boldsymbol{\beta}$ *but not on the choice of* J.

Let $\Theta_{\boldsymbol{\beta}} : \mathfrak{g} \to \mathfrak{g}^*$ be the symplectic cocycle of the Lie algebra $\mathfrak{g}$ associated to the symplectic 1-cocycle $\theta_{\boldsymbol{\beta}}$ of the Lie group G (see, for example, [18]). Its expression is

$$\Theta_{\boldsymbol{\beta}}(X) = T_e\theta_{\boldsymbol{\beta}}(X) = \Theta(X) - \mathrm{ad}^*_X E_J(\boldsymbol{\beta})\,.$$

The map $\Theta_{\boldsymbol{\beta}}$ is the unique symplectic 1-cocycle of the Lie algebra $\mathfrak{g}$ which is cohomologous to Θ and satisfies the equality

$$\Theta_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = 0\,.$$

Let X and Y be two elements in $\mathfrak{g}$, considered as two elements of $T_{\boldsymbol{\beta}}\Omega$, in other words as two vectors tangent to Ω at its point $\boldsymbol{\beta}$. Let us moreover assume that X is tangent to the adjoint orbit of $\boldsymbol{\beta}$ at its point $\boldsymbol{\beta}$. There exists $X_1 \in \mathfrak{g}$ such that $X = [\boldsymbol{\beta}, X_1]$. When evaluated on the pair of tangent vectors (X, Y), the Riemannian metric Γ can be expressed as

$$\Gamma(\boldsymbol{\beta})(X, Y) = \bigl\langle \Theta_{\boldsymbol{\beta}}(X_1), Y \bigr\rangle\,.$$

If Y too is tangent to the adjoint orbit of $\boldsymbol{\beta}$ at its point $\boldsymbol{\beta}$, there exists $Y_1 \in \mathfrak{g}$ such that $Y = [\boldsymbol{\beta}, Y_1]$, and we have the two equalities, which express the Riemannian metric induced by Γ on the adjoint orbit of $\boldsymbol{\beta}$,

$$\Gamma(\boldsymbol{\beta})(X, Y) = \bigl\langle \Theta_{\boldsymbol{\beta}}(X_1), Y \bigr\rangle = \bigl\langle \Theta_{\boldsymbol{\beta}}(Y_1), X \bigr\rangle\,.$$

Proof: Since Θ, being a symplectic cocycle, is skew-symmetric, we have, for each $X \in \mathfrak{g}$, $\bigl\langle \Theta(\boldsymbol{\beta}), X \bigr\rangle = -\bigl\langle \Theta(X), \boldsymbol{\beta} \bigr\rangle$. Using the equalities proven in Proposition3.3.3, we obtain

$$\begin{aligned}\bigl\langle \Theta_{\boldsymbol{\beta}}(\boldsymbol{\beta}), X \bigr\rangle &= \bigl\langle \Theta(\boldsymbol{\beta}), X \bigr\rangle - \bigl\langle \mathrm{ad}^*_{\boldsymbol{\beta}} E_J(\boldsymbol{\beta}), X \bigr\rangle = -\bigl\langle \Theta(X), \boldsymbol{\beta} \bigr\rangle - \bigl\langle E_J(\boldsymbol{\beta}), [\boldsymbol{\beta}, X] \bigr\rangle \\ &= -\bigl\langle E_J(\boldsymbol{\beta}), [X, \boldsymbol{\beta}] \bigr\rangle - \bigl\langle E_J(\boldsymbol{\beta}), [\boldsymbol{\beta}, X] \bigr\rangle = 0\,.\end{aligned}$$

Other statements about $J_{\boldsymbol{\beta}}$, $\theta_{\boldsymbol{\beta}}$ and $\Theta_{\boldsymbol{\beta}}$ easily follow from well known properties of moment maps of Hamiltonian actions (see for example [18]).

Using Theorem 3.2.4 and Proposition 3.3.3, we obtain, for all $\boldsymbol{\beta} \in \Omega$, X_1 and $Y \in \mathfrak{g}$, with $X = [X_1, \boldsymbol{\beta}]$,

$$\Gamma(\boldsymbol{\beta})\bigl([X_1, \boldsymbol{\beta}], Y\bigr) = -\bigl\langle DE_J(\boldsymbol{\beta})\bigl([X_1, \boldsymbol{\beta}]\bigr), Y \bigr\rangle = \bigl\langle \mathrm{ad}^*_{[X_1, \boldsymbol{\beta}]} E_J(\boldsymbol{\beta}) + \Theta(X_1), Y \bigr\rangle\,.$$

According to Proposition 3.2.6, the bilinear form Γ does not depend on the choice of the moment map J, so we can replace J by $J_{\boldsymbol{\beta}}$ in the right hand side of the above equality. Of course we have to replace too E_J by $E_{J_{\boldsymbol{\beta}}}$ and Θ by $\Theta_{\boldsymbol{\beta}}$. The map $J_{\boldsymbol{\beta}}$ was chosen so that$E_{J_{\boldsymbol{\beta}}}(\boldsymbol{\beta}) = 0$, so we obtain

$$\Gamma(\boldsymbol{\beta})\bigl([X_1, \boldsymbol{\beta}], Y\bigr) = \bigl\langle \Theta_{\boldsymbol{\beta}}(X_1), Y \bigr\rangle\,.$$

When we both have $X = [X_1, \boldsymbol{\beta}]$ and $Y = [Y_1, \boldsymbol{\beta}]$, with X_1 and $Y_1 \in \mathfrak{g}$, we can exchange the parts played by X and Y and write

$$\begin{aligned}\Gamma(\boldsymbol{\beta})(X, Y) &= \Gamma(\boldsymbol{\beta})\bigl([X_1, \boldsymbol{\beta}], [Y_1, \boldsymbol{\beta}]\bigr) = \Gamma(\boldsymbol{\beta})\bigl([Y_1, \boldsymbol{\beta}], [X_1, \boldsymbol{\beta}]\bigr)\\ &= \Gamma(\boldsymbol{\beta})\bigl([Y_1, \boldsymbol{\beta}], X\bigr) = \bigl\langle \Theta_{\boldsymbol{\beta}}(Y_1), X \bigr\rangle\,.\end{aligned}$$

■

4. Final comments

Gibbs states built with a Hamiltonian which does not depend on time appear as good models of physicists' states of thermodynamic equilibrium : they are invariant by the flow of the associated Hamiltonian vector field (proposition 2.3.3), and the properties of the real parameter β which indexes the set of Gibbs states are in good agreement with those of the inverse of an absolute temperature (subsection 2.3.5). However, the flow of a Hamiltonian vector field whose Hamiltonian does not depend on time cannot describe the evolution with time of a statistical state towards the corresponding state of thermodynamic equilibrium, since the state's entropy is invariant by this flow (proposition 2.2.10).

Properties of Gibbs states built with the moment map of the Hamiltonian action of a Lie group are very similar to those of Gibbs states built with a Hamiltonian which does not depend on time as the only conserved quantity : the set of generalized temperatures is an open convex subset of the Lie algebra, the partition function, the mean value of the moment map and the entopy are smooth functions of the generalized temperature. When the considered Lie group is not commutative, remarkable new equivariance properties with respect to the adjoint action of this Lie group appear, which did not exist for Gibb states built with the Hamiltonian as the only conserved quantity (subsection 3.3). Generalized temperatures only exist when there exist a non-empty subset Ω of elements $\boldsymbol{\beta}$ in the Lie algebra of the considered Lie group for which the integral which defines the partition function is normally convergent (definition 3.1.2). When the considered symplectic manifold is not compact, one may think that the restrictions imposed by this condition increase with the dimension of the considered Lie group. Our next paper will present examples of Gibbs states for the Hamiltonian action of a non-commutative Lie group on a symplectic manifold, even when the considered symplectic manifold is not compact, and examples in which no Gibbs state can exist, the set of generalized temperatures being empty.

Acknowledgements

Although I studied Jean-Marie Souriau's book [27] in my youth, my knowledge of his works in statistical mechanics was rather superficial. I owe my interest in Gibbs states to Frédéric Barbaresco, who led me to look again at this book more in depth.

Roger Balian and Alain Chenciner were kind enough to look at this work. Their numerous helpful remarks and constructive criticisms, their respective helps to better understand the foundations of quantum statistical mechanics on one hand and Shannon's paper [25] on the other hand, were invaluable.

I benefited of very fruitful exchanges of ideas with Géry de Saxcé during our common work on a paper in preparation.

To all of them, and to my colleagues and friends I encountered regularly in various seminars before the beginning of the present dark period of sanitary confinement, I address my warmest thanks.

References

[1] Balian R., *Information in Statistical Physics*, Studies in History and Philosophy of Modern physics, **36** (2005), pp.323–353.

[2] Balian R., *François Massieu et les Potentiels Thermodynamiques*, Évolution des disciplines et histoire des découvertes, Académie des Sciences, Avril 2015.

[3] Barbaresco F., *Koszul Information Geometry and Souriau Geometric Temperature/Capacity of Lie Group Thermodynamics*, Entropy, vol. 16, 2014, pp. 4521-4565. Published in the book *Information, Entropy and Their Geometric Structures*, MDPI Publisher, September 2015.

[4] Barbaresco F., *Symplectic Structure of Information Geometry: Fisher Metric and Euler-Poincaré Equation of Souriau Lie Group thermodynamics*. In *Geometric Science of Information, Second International Conference GSI 2015 Proceedings*, (Franck Nielsen and Frédéric Barbaresco, editors), Lecture Notes in Computer Science vol. 9389, Springer 2015, pp. 529–540.

[5] Barbaresco F., *Geometric Theory of Heat from Souriau Lie Groups thermodynamics and Koszul Hessian Geometry: Applications in Information Geometry for Exponential Families*, in the Special Issue "Differential Geometrical Theory of Statistics", MDPI, Entropy **2016**, 18(11), 386; doi:10.3390/e18110386.

[6] Barbaresco F., *Lie Group Statistics and Lie Group Machine Learning Based on Souriau Lie Groups Thermodynamics and Koszul-Souriau-Fisher Metric: New Entropy Definition as Casimir Invariant Function in Coadjoint Representation*, Entropy 2020, 22, 642; doi:10.3390/e22060642.

[7] Barbaresco F., *Souriau Entropy Based on Symplectic Model of Statistical Physics: three Jean-Marie Souriau's Seminal Papers on Lie Groups Thermodynamics*, Preprint, partial English translation of the papers [28–30] by Jean-Marie Souriau. `https://www.academia.edu/44444245/Souriau_Entropy_based_on_Symplectic_Model_of_Statistical_Physics_three_Jean_Marie_Souriaus_seminal_papers_on_Lie_Groups_Thermodynamics`.

[8] Barbaresco F., Gay-Balmaz F., *Lie Group Cohomology and (Multi)Symplectic Integrators: New Geometric Tools for Lie Group Machine Learning Based on Souriau Geometric statistical mechanics*, Entropy 2020, 22, 498; doi:10.3390/e22050498.

[9] Boltzmann L., *Leçons sur la théorie des gaz*, Gauthier-Villars, Paris, 1902–1905, reprinted by Éditions Jacques Gabay, Paris, 1987. The second part can be freely downloaded at `http://iris.univ-lille1.fr/handle/1908/1523`.

[10] Chenciner A., *La force d'une idée simple, hommage à Claude Shannon à l'occasion du centenaire de sa naissance*, Gazette des mathématiciens, Société mathématique de France, n. 152, avril 2017, pp. 16–22.

[11] Fréchet M., *Sur l'extension de certaines évaluations statistiques au cas de petits échantillons*, Revue de l'Institut International de Statistique / Review of the International Statistical Institute, **11**, No. 3/4 (1943), pp. 182-205.

[12] Gibbs W., *Elementary Principles in Statistical Mechanics, developed with Especial Reference to the Rational Foundation of Thermodynamics*, New York: Charles Scribner's sons, London: Edward Arnold, 1902. The camera-quality files for this public-domain ebook may be downloaded *gratis* at `www.gutenberg.org/ebooks/50992`.

[13] Jaynes T., *Information Theory and Statistical Mechanics*, Phys. Rev. vol. 106, n. 4 (1957), pp. 620–630.

[14] Jaynes, E. T., *Information Theory and Statistical Mechanics II*, Phys. Rev. vol. 108, n. 2 (1957), pp. 171–190.

[15] Jaynes T., *Information Theory and Statistical Mechanics*, in *Statistical Physics*, Brandeis Lectures in Theoretical Physics, volume 3, K. Ford (editor), Benjamin, New York, 1963, pp. 181–218.

[16] Jaynes T., *Prior Probabilities*, IEEE Transactions On Systems Science and Cybernetics, vol. sec-4, no. 3, 1968, pp. 227–241.

[17] Mackey W., *The Mathematical Foundations of Quantum mechanics*, W. A. Benjamin, Inc., New York, 1963.

[18] Marle C.-M., *Géométrie Symplectique et Géométrie de Poisson*, Calvage & Mounet, Paris, 2018.

[19] Massieu F., *Sur les Fonctions Caractéristiques des Divers Fluides*, C. R. Acad. Sci. Paris vol. 69, 1869, pp. 858–862.

[20] Massieu F., *Addition au précédent Mémoire sur les Fonctions Caractéristiques*, C. R. Acad. Sci. Paris vol. 69, 1869, pp. 1057–1061.

[21] Massieu F., *Thermodynamique. Mémoire sur les Fonctions Caractéristiques des Divers Fluides et sur la Théorie des Vapeurs*, mémoires présentés par divers savants à l'Académie des Sciences de l'Institut National de France, XXII, n. 2, 1876, pp. 1–92.

[22] Matheron G., *Éléments pour une théorie des milieux poreux*, Masson, Paris, 1967.

[23] de Saxcé G., *Entropy and Structure for the Thermodynamic Systems*, in *Geometric Science of Information, Second International Conference GSI 2015 Proceedings*, (Franck Nielsen and Frédéric Barbaresco, editors), Lecture Notes in Computer Science vol. 9389, Springer 2015, pp. 519–528.

[24] de Saxcé, G., *Link Between Lie Group Statistical Mechanics and Thermodynamics of Continua*. In the special Issue "Differential Geometrical Theory of Statistics", MDPI, Entropy, **2016**, 18, 254; doi:10.3390/e18070254.

[25] Shannon C., *A Mathematical Theory of Communication*, The Bell System Technical Journal, vol. 27, pp. 379–423 and 623–656, July and October 1948. This paper can be freely downloaded at https://web.archive.org/web/19980715013250/http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf.

[26] Souriau J.-M., *Définition Covariante des Équilibres Thermodynamiques*, Supplemento al Nuovo cimento vol. IV n.1, 1966, pp. 203–216.

[27] Souriau J.-M., *Structure des Systèmes Dynamiques*, Dunod, Paris, 1969. English translation: *Structure of Dynamical Systems, a Symplectic View of Physics*, translated by C. H. Cushman-de Vries, translation editors R. H. Cushman, G. M. Tuynman, Progress in Mathematics volume 149, Birkhäuser Boston, 1997.

[28] Souriau J.-M., *Mécanique Statistique, Groupes de Lie et Cosmologie*, Colloques internationaux du CNRS numéro 237 *Géométrie Symplectique et Physique Mathématique*, 1974, pp. 59–113.

[29] Souriau J.-M., *Géométrie Symplectique et Physique Mathématique*, deux conférences de Jean-Marie Souriau, Colloquium de la Société Mathématique de France, 19 février et 12 novembre 1975.

[30] Souriau J.-M., *Mécanique Classique et Géométrie Symplectique*, preprint, Université de Provence et Centre de Physique Théorique, 1984.

[31] Wikipedia article *Limiting density of discrete points*, https://en.wikipedia.org/wiki/Limiting_density_of_discrete_points.

Charles-Michel Marle
Honorary Professor, retired from the Université Pierre et Marie Curie (today Sorbonne Université), Paris, France.

Home Address: 27, avenue du 11 novembre 1918, 92190 Meudon, France

E-mail address: cmm1934@orange.fr

Other E-mail address: charles-michel.marle@math.cnrs.fr